



Title Vocabulary acquisition through extensive
reading of unsimplified English material in a
Saudi Arabian tertiary context

Name Hussain Ziya Zaeem Sivardeen

This is a digitised version of a dissertation submitted to the University of Bedfordshire.

It is available to view only.

This item is subject to copyright.



**VOCABULARY ACQUISITION THROUGH EXTENSIVE
READING OF UNSIMPLIFIED ENGLISH MATERIAL IN
A SAUDI ARABIAN TERTIARY CONTEXT**

Hussain Ziya Zaeem Sivardeen

This is a digitised version of a dissertation submitted to the
University of Bedfordshire.

It is available to view only.

This item is subject to copyright.

**VOCABULARY ACQUISITION THROUGH EXTENSIVE
READING OF UNSIMPLIFIED ENGLISH MATERIAL IN
A SAUDI ARABIAN TERTIARY CONTEXT**

HUSSAIN ZIYA ZAEEM SIVARDEEN

A thesis submitted to the University of Bedfordshire in fulfilment
of
the requirements for the degree of Doctor of Philosophy

University of Bedfordshire
Centre for Research in English Language Learning and
Assessment
(CRELLA)

October 2016

VOCABULARY ACQUISITION THROUGH EXTENSIVE READING OF UNSIMPLIFIED ENGLISH MATERIAL IN A SAUDI ARABIAN TERTIARY CONTEXT

HUSSAIN ZIYA ZAEEM SIVARDEEN

ABSTRACT

A number of studies seem to indicate that Extensive Reading (ER) enhances language proficiency in general and vocabulary knowledge in particular. However none of those surveyed focused on poorly motivated lower level language learners in a normal course of study that incorporated ER of unsimplified material, and few addressed partial vocabulary gain and the other factors that could have affected this.

In this study the participants were young adult male Saudi Arabians, who were elementary English language learners and who were not used to reading for pleasure even in Arabic. Tests were used to gauge their partial vocabulary acquisition when they engaged in ER involving an open choice of simplified and unsimplified English material during a regular English preparatory course of just one semester. If ER could be linked to improved vocabulary acquisition in such difficult but natural conditions, the case for ER's wider implementation could be strengthened. The study also used diaries, surveys and interviews to delve into the reading habits, academic background and cultural context that could also explain any observed vocabulary gain. The results showed there were few statistically significant partial vocabulary gains after the ER programme, and although the experimental cohort that received the ER programme showed greater gains than the control in most of these cases, the difference was not large in terms of the absolute number of words. The diaries, surveys and interviews gave a rich profile of the participants with a level of detail that surpassed that of any other surveyed study, offering several possible reasons for their modest vocabulary gains and yielding unanticipated findings, such as an exposition of their motivations for academic study.

It is concluded that a greater awareness of the learners' context is essential when applying ER, which in this particular context leads to the view that pure ER may not be suitable for vocabulary learning, and instead a significantly modified version could be more appropriate. These modifications include selecting appropriate books for the students and adding explicit vocabulary learning activities, and they even include linking the reading to the students' final grade. Further research is needed to gauge the effectiveness of such modifications, while a more inductive approach is also important when investigating ER and vocabulary acquisition to give the opportunity for unforeseen results to emerge.

AUTHOR’S DECLARATION

I, Hussain Ziya Zaeem Sivardeen, declare that this thesis and the work presented in it are my own and that they have been generated by me as the result of my own original research.

“VOCABULARY ACQUISITION THROUGH EXTENSIVE READING OF
UNSIMPLIFIED ENGLISH MATERIAL IN A SAUDI ARABIAN TERTIARY
CONTEXT”

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have cited the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission.

Signature for hard copy:



ACKNOWLEDGEMENTS

I would firstly like to express my sincere thanks to the administrators, colleagues and students who supported this study by giving formal permission for its on-site activities, by giving detailed feedback during the pilot stages, by assisting with the control cohort, and by participating as subjects.

I am also very grateful to the authors of the Vocabulary Size Test and the Vocabulary Levels Test for giving formal permission for their use and modification.

In addition, I must show my great appreciation to the Supervisory Team at the University of Bedfordshire. I can say without a doubt that I have never before received such outstanding constructive advice and such excellent motivation during any academic endeavour. If I could ever emulate this brilliant mentoring to my own students, then the Supervisory Team at the University of Bedfordshire would have to take the credit.

Furthermore, I am immensely thankful to my wife and children who were patient with me during the time I was preoccupied with this study.

Finally, of all the people in this world, the greatest acknowledgement has to go to my parents. I will never be able to show enough gratitude to them for their pivotal role in making me what I am today.

CONTENTS

ABSTRACT	p.iii
AUTHOR'S DECLARATION	p.iv
AKNOWLEDGEMENTS	p.v
LIST OF TABLES	p.ix
LIST OF FIGURES	p.ix
LIST OF APPENDICES	p.x
 CHAPTER-1	 p.1
INTRODUCTION	
 CHAPTER-2	 p.5
LITERATURE REVIEW	
2.1 EXTENSIVE READING	p.5
2.2 INTENSIVE READING	p.6
2.3 ER AND READING	p.7
2.4 ER AND OTHER ASPECTS OF LANGUAGE ACQUISITION	p.8
2.5 LIMITATIONS WITH ER	p.13
2.6 ER AND VOCABULARY ACQUISITION	p.16
2.7 LIMITATIONS IN PREVIOUS STUDIES ON ER & VOCABULARY	p.19
2.7.1 Limitations in Studies using Graded Readers	p.20
2.7.2 Limitations in Studies using Unsimplified Material	p.26
2.7.3 Limitations in Purely Experimental Studies placed in Non-Teaching or Non-ER Contexts	p.30
2.7.4 Limitations in Studies that did not consider Other Sources of English Exposure	p.32
2.7.5 Limited Consideration of Partial Vocabulary Acquisition	p.33
2.7.6 Limited Research of Lower-Level Saudi University Students	p.34
2.8 RESEARCH QUESTIONS	p.36
2.8.1 Definitions of Key Concepts in the Research Questions	p.36
2.8.2 Distinctive Aspects of this Study	p.38
 CHAPTER-3	 p.41
METHODOLOGY	
3.1 METHODOLOGY OVERVIEW	p.42
3.1.1 Development, Context and Participants	p.42
3.1.2 Overview of Procedure during the Main Study	p.45
3.1.3 Research Design	p.50
3.2 METHODOLOGY FOR THE ER PROGRAMME	p.54
3.2.1 Incorporation of the ER Programme into the Normal Course of Study	p.57
3.2.2 Open Choice of Reading Materials	p.57
3.2.3 Sustained Silent Reading Sessions	p.61
3.2.4 Encouragement to Read Outside the Class	p.65
3.2.5 Duration of ER Programme	p.66
3.2.6 Control Cohort	p.66
3.3 METHODOLOGY FOR THE PLACEMENT TEST	p.69

CONTENTS (cont.)

CHAPTER-3

METHODOLOGY (cont.)

3.4	METHODOLOGY FOR TESTING VOCABULARY GAIN	p.72
3.4.1	Testing Specific Aspects of Word Knowledge	p.73
3.4.2	Testing for Partial Vocabulary Acquisition	p.75
3.4.3	Overview of Vocabulary Tests Used	p.76
3.4.4	Overview of Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning)	p.79
3.4.4.1	<i>Details of Test-1 (Spelling)</i>	p.82
3.4.4.2	<i>Details of Test-2 (Translation)</i>	p.84
3.4.4.3	<i>Details of Test-3 (Meaning)</i>	p.85
3.4.5	Test-4 (Meaning)	p.86
3.4.6	Administration of Tests	p.89
3.4.7	Statistical Analysis of the Main Pilot Test Results	p.90
3.4.7.1	<i>Internal Consistency</i>	p.96
3.4.7.2	<i>Results for the Different Aspects of Word Knowledge</i>	p.96
3.4.7.3	<i>Results for the Different Word Frequency Levels</i>	p.97
3.4.7.4	<i>Results involving Levels of Certainty</i>	p.98
3.4.7.5	<i>Distribution of Data</i>	p.98
3.4.7.6	<i>Vocabulary Gain between Pretests and Posttests</i>	p.99
3.4.7.7	<i>Summary</i>	p.100
3.5	METHODOLOGY FOR THE END-OF-SEMESTER EXAM	p.101
3.6	METHODOLOGY FOR THE READING DIARIES	p.103
3.6.1	Development of Items	p.104
3.6.2	Layout of Content	p.107
3.6.3	Limitations	p.108
3.6.4	Procedure	p.108
3.7	METHODOLOGY FOR THE SURVEY OF EXPOSURE TO ENGLISH LEARNING	p.111
3.7.1	Development of Items	p.112
3.7.2	Layout of Content	p.113
3.7.3	Limitations	p.114
3.7.4	Procedure	p.115
3.8	METHODOLOGY FOR THE SUPPLEMENTARY QUESTIONING	p.117
3.9	METHODOLOGY FOR THE IN-DEPTH INTERVIEWS	p.118
3.9.1	Participants	p.119
3.9.2	Procedure	p.120
3.9.2.1	<i>Interview Schedule</i>	p.124
3.9.3	Compensating for Potential Limitations	p.127
3.10	METHODOLOGY FOR THE RESEARCH JOURNAL	p.129

CONTENTS (cont.)

CHAPTER-4	p.131
RESULTS	
4.1 PLACEMENT TEST RESULTS	p.134
4.1.1 Results	p.134
4.2 VOCABULARY TESTS RESULTS	p.135
4.2.1 Internal Consistency	p.141
4.2.2 Results for the Different Aspects of Word Knowledge	p.142
4.2.3 Results for the Different Word Frequency Levels	p.142
4.2.4 Results involving Levels of Certainty	p.143
4.2.5 Distribution of Data	p.144
4.2.6 Vocabulary Gain between Pretests and Posttests	p.145
4.2.7 Summary	p.147
4.3 END-OF-SEMESTER EXAM RESULTS	p.149
4.3.1 Results	p.149
4.4 READING DIARIES RESULTS	p.151
4.4.1 Results	p.154
4.4.2 Summary	p.156
4.5 SURVEY OF EXPOSURE TO ENGLISH LEARNING RESULTS	p.156
4.5.1 Results	p.159
4.5.2 Summary	p.161
4.6 SUPPLEMENTARY QUESTIONING RESULTS	p.162
4.6.1 Results	p.162
4.7 IN-DEPTH INTERVIEWS RESULTS	p.164
4.7.1 Results	p.177
4.7.2 Summary	p.192
CHAPTER-5	p.195
DISCUSSION	
5.1 DISCUSSION OF RESULTS RELEVANT TO RESEARCH QUESTION 1	p.195
5.1.1 Summary: Response to Research Question 1	p.199
5.2 DISCUSSION OF RESULTS RELEVANT TO RESEARCH QUESTION 2	p.201
5.2.1 Discussion: Reading Diaries	p.201
5.2.2 Discussion: Survey of Exposure to English Learning	p.204
5.2.3 Discussion: In-Depth Interviews	p.206
5.2.4 Summary: Response to Research Question 2	p.212
5.3 DISCUSSION OF RESULTS THAT WENT BEYOND THE RESEARCH QUESTIONS	p.214
5.3.1 Summary	p.220
CHAPTER-6	p.223
CONCLUSIONS	
6.1 CONTRIBUTION TO KNOWLEDGE	p.223
6.1.1 Summary	p.230
6.2 IMPLICATIONS FOR RESEARCH & PRACTICE	p.231
6.2.1 Summary & Final Thoughts	p.237
BIBLIOGRAPHY	p.239

CONTENTS (cont.)

LIST OF TABLES

Table-1	A Comparison of ER and IR based on the Reviewed Literature in Section-2.1 & Section-2.2	Section-2.2	p.7
Table-2	A Summary of the Stages of this Study	Section-3.1.1	p.42
Table-3	Number of Participants in this Study and other Surveyed Studies on Reading and Vocabulary	Section-3.1.1	p.43
Table-4	Data Collection Instruments during the Main Study	Section-3.1.2	p.47
Table-5	Timeline for the Procedure during the Main Study Semester	Section-3.1.2	p.48
Table-6	Words Tested in this Study and other Surveyed Studies on Vocabulary Gain through Reading	Section-3.1.2	p.48
Table-7	Timeline for the ER Programme of the Main Study Experimental Cohort during its Teaching Semester	Section-3.2	p.56
Table-8	Sustained Silent Reading (SSR) in this Study and other Surveyed ER Studies	Section-3.2.3	p.64
Table-9	ER Programme Duration in this Study and other Surveyed ER Studies	Section-3.2.5	p.66
Table-10	Facets of Word Knowledge (Nation, 2004:22)	Section-3.4.1	p.74
Table-11	Summary of the Vocabulary Tests Used in This Study	Section-3.4.3	p.77
Table-12	Brief Summary of the Vocabulary Tests Used in This Study	Section-4.2	p.135

LIST OF FIGURES

Figure-1	The First Few Questions of the Revised Versions of Test-1 (Spelling) & Test-2 (Open Translation) Combined [See Appendix-6 for the full version]	Section-3.4.4.1	p.83
Figure-2	The First Few Questions of the Revised Version of Test-3 (Meaning) [See Appendix-8 for the full version]	Section-3.4.4.3	p.86
Figure-3	The First Few Questions of the Revised Versions of Test-4 (Meaning) [See Appendix-10 for the full version]	Section-3.4.5	p.88
Figure-4	Example Entries for the First Three Items of Test-3 (Meaning) [See Appendix-8]	Section-3.4.7	p.93
Figure-5	Summary of Vocabulary Test Results during the Main Pilot [See Appendix-15]	Section-3.4.7	p.95
Figure-6	Summary of Vocabulary Test Results during the Main Study comparing Experimental & Control Cohorts (change in medians given to 1 d.p.) [See Appendix-18]	Section-4.2	p.138
Figure-7	Summary of Vocabulary Test Results during the Main Study - Experimental Cohort [See Appendix-18]	Section-4.2	p.139
Figure-8	Summary of Vocabulary Test Results during the Main Study – Control Cohort [See Appendix-18]	Section-4.2	p.140
Figure-9	Layout of the Excel Spreadsheet used in Phase-2 to assign Initial Codes to Extracts, showing the two example extracts detailed in the main text	Section-4.7	p.170
Figure-10	Results of the In-Depth Interviews: (A) Possible Factors for the Modest Vocabulary Gains during the ER Programme	Section-4.7.1	p.179
Figure-11	Results of the In-Depth Interviews: (B) Extrinsically Motivated Students	Section-4.7.1	p.184
Figure-12	Results of the In-Depth Interviews: (C) Future Implementation of ER	Section-4.7.1	p.187

CONTENTS (cont.)

LIST OF APPENDICES

Appendix-1	The Original Words in the Vocabulary Size Test (Nation, 2009) up to the 10000-word level [see Section-3.4.4]	p.259
Appendix-2	The Chosen Words for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) [see Section-3.4.4 for the selection criteria]	p.261
Appendix-3	Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2]	p.262
Appendix-4	Test-3 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.4.3]	p.266
Appendix-5	Summary of Specific Revisions made to Test-1 v.1 (Spelling) & Test-3 v.1 (Meaning) in light of the feedback during the Stage-1 Initial Pilot Study [see Section-3.4.4.3]	p.269
Appendix-6	Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1]	p.270
Appendix-7	Final List of Acceptable Translated Answers for Test-2 (Open Translation) during the Stage-3 Main Study [see Section-3.4.4.2]	p.274
Appendix-8	Test-3 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.3]	p.276
Appendix-9	Test-4 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.5]	p.279
Appendix-10	Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.5]	p.282
Appendix-11	Daily Reading Diary Sheet - Stage-1 Initial Pilot Study [see Section-3.6.1]	p.286
Appendix-12	Weekly Reading Diary Sheet - Stage-3 Main Study [see Section-3.6.1, Section-3.6.2 & Section-4.4]	p.287
Appendix-13	Survey of Exposure to English Learning - Stage-3 Main Study [see Section-3.7.1]	p.288
Appendix-14	Results of English Placement Test [see Section-4.1]	p.290
Appendix-15	Results of the Stage-2 Main Pilot Vocabulary Tests [see Section-3.4.7]	p.292
Appendix-16	Comparing Quantitative Data of the Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49) [see Section-3.9.1]	p.294
Appendix-17	In-Depth Interview Schedule [see Section-3.9.2.1]	p.297
Appendix-18	Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]	p.300
Appendix-19	Results of End-of-Semester English Course Exam [see Section-4.3]	p.305
Appendix-20	Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4]	p.307
Appendix-21	Results of the Survey of Exposure to English Learning [see Section-4.5.1]	p.314
Appendix-22	A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7]	p.320
Appendix-23	Initial Codes generated during Phase-2 of the Interview Analysis [see Section-4.7]	p.337
Appendix-24	Main Codes & Sub-Codes generated during Phase-3 of the Interview Analysis [see Section-4.7]	p.338
Appendix-25	Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7]	p.342
Appendix-26	Narratives and Themes refined during Phase-5 & Phase-6 of the Interview Analysis [see Section-4.7]	p.346

CHAPTER-1

INTRODUCTION

English language teaching in Saudi Arabia has increased dramatically in recent years since the medium of instruction for all higher secular education switched to English. Teachers - like myself - were tasked with raising the level of our students' English from almost nothing to that of being able to study their degree subjects in English, all within a year or a semester depending on the institution. In my particular context, the students were young men, typically in their early 20s, coming to the capital Riyadh from all over Saudi Arabia, including the small towns and villages, to study bachelor degrees in different technical specializations. They were from the poorer levels of society, their prior education was in Arabic at state schools, and they were lower-level in English ability. They were apathetic to reading, even in their own language, with most not reading at all unless forced to do so for academic exams. All of this created a context that can be described, in the very least, as challenging.

Reading is considered 'one of the most important skills for academic success' (Schmitt et al., 2011:26), and Extensive Reading (ER) can develop this. ER has been defined as reading large quantities of a target language (Hill, 2008:186) for enjoyment and information (Bamford, 1984:219), and not for any specific language learning purpose or formal study (Al-Homoud & Schmitt, 2009:383; Day & Bamford, 2002:138), with its incorporation into English for Academic Purposes (EAP) courses being specifically recommended (Benson, 1991:75; Carrell & Carson, 1998:47).

Consequently, ER could have an important role to play in EAP courses in Saudi Arabia. Unfortunately, as mentioned above, I have found the students in this context to be averse to reading for pleasure in their own language, let alone in English. They readily report that they may read about football in the newspapers or the internet, but little else unless forced to by academic studies or work requirements. However, this is actually part of the reason why this context is so thought-provoking, especially given that this

attitude to reading for pleasure is not only restricted to my Saudi students. In England, a generation ago at school, it was true that a few friends were immersed in novels. They would read the adventures of *The Famous Five*, *Sherlock Holmes*, *Hercule Poirot*, *Charlie and the Chocolate Factory*, etc. and they were undeniably amongst the best at school in English and other subjects.

But what about the rest, who formed the overwhelming majority? For them it was television: ITV's *The Famous Five*, Granada's *Sherlock Holmes*, LWT's *Agatha Christie's Poirot*, and the film *Willy Wonka & the Chocolate Factory*. All were watched, but few were read. And now this generation of young learners seems to have even more distractions. They seem preoccupied with computers, laptops, tablets and smart phones, spending all their time accessing films, video clips, social media and chats via the internet. Reading large amounts of English prose for pleasure appears to be strange even in England (as also reported by Brown, 2013), the birth place of English. Saudi Arabia is certainly going to be a more demanding environment. If reading for pleasure, and hence overall academic language proficiency can be greatly improved here, then that would definitely be a huge achievement.

In addition, being both an active teaching practitioner and an academic intellectual, my primary concerns relate to research that is linked to real teaching contexts with real students and real constraints. Hence I view research and practice as intertwined and mutually inclusive, and so I am keen to investigate ER in the actual classroom environment, because any findings would surely be of immediate benefit to frontline teachers, as well as theorists who wish to develop their ideas based on data from the field.

My personal professional background also colours the methodology adopted for this investigation. As a teacher, I use both summative and formative assessments of my students, recognizing that both are important and complementary in providing a more complete picture of their progress. Furthermore, I can say that I personally do not need tests with percentage scores to know which of my students are the best or the worst, because I am immersed with them in their learning every lesson and every day, enabling

me to gauge their level continuously, to a deeper and richer degree than any summative test could. However, many other stakeholders, such as parents and college administrators are not constantly present in the classroom, and they do require percentage scores, which they consider to be more objective and defensible than the teacher's subjective views, no matter how well founded they may be. As a result, my study will involve collecting both quantitative and qualitative data. This reflects my view that they are equally significant, and this makes the investigation more accessible to other teachers and researchers who hold a wide variety of educational and research beliefs, adding to the overall body of knowledge that is relevant to all stakeholders in language learning.

Hence this is the broad background of the context I am in, from which emerges my study: the implementation of a ER programme that is incorporated into a normal teaching course, to investigate its effects on the language proficiency of Saudi university students, who read little even in L1. I believe this is an 'intrinsically interesting' study (McDonough & McDonough, 1997:84) because if ER can be linked to improved proficiency in such challenging but natural conditions, then the case for ER's wider implementation could be strengthened in Saudi Arabia and beyond.

The following chapter of this thesis is Chapter-2 (Literature Review), which gives a detailed survey of previous relevant studies. It explains the development of my exploration into ER and vocabulary acquisition, culminating in precisely defined research questions, noting that relatively few studies were found that dealt with this context of poorly motivated, young adult male Saudi university students, and even fewer were found that dealt with the other distinctive aspects of my study. These distinctive aspects are also detailed at the end of the chapter, and they include the implementation of a relatively short ER course of just one semester that encourages reading unsimplified material from all sources (not just story books with deliberately simplified English); the use of partial vocabulary acquisition as a convenient gauge to measure improved language proficiency during the ER course, noting the key role of vocabulary in language, and noting that vocabulary is not gained in an all-or-nothing way; and also the study of other factors that may be associated with vocabulary

acquisition, noting that factors outside the classroom may play a significant role in language learning.

This is followed by Chapter-3 (Methodology), which details what I did to investigate the research questions. It describes the pilot studies that were used to formulate the ER programme and the array of data collection instruments (tests, diary sheets, surveys, interviews & supplementary questions), the rationale for using them, their development, and their administration. Then in Chapter-4 (Results), the findings are presented, as well as the methodology for how the raw data was analysed to yield those findings. Chapter-5 (Discussion) scrutinizes and debates these results, and Chapter-6 (Conclusions) closes with what can be learned from the project in its entirety and in its specifics.

With this, Chapter-1 (Introduction) is concluded, and so now begins the rest of the thesis starting with Chapter-2 (Literature Review).

CHAPTER-2

LITERATURE REVIEW

The act of reading has been described as ‘a number of *interactive processes* involving the reader and the text...’ where readers ‘...use their knowledge of the world, the topic, [and] the language, to interact with the text to *create, construct, or build* meaning’ (Day, 2007:19). It is also an activity that can combine both learning and enjoyment (Nation, 2001:1), and as with any enjoyable method of learning, it holds an important position in language acquisition.

2.1 EXTENSIVE READING

Throughout the past 60 years, Extensive Reading (ER) has been given different names and has been described in different ways (Day, 2015:294). These can be summarized as involving reading large quantities of a target language (Yamashita, 2015:169) in order to ‘flood’ learners with... L2 input’ (Hafiz & Tudor, 1989:5) so that although some ER may be done in class, most needs to be done outside (Brown, 2009:241).

In order to encourage this large amount of reading, the material should to be ‘pleasurable’ to read (Pigada & Schmitt, 2006:2), and so ER is for enjoyment and information (Bamford, 1984:219). The material should also be ‘at a comfortable level for the learner’ (Brown, 2009:241) and within the reader’s ‘linguistic competence’ (Pigada & Schmitt, 2006:2). In order to achieve both these characteristics of enjoyment and ease, it is recommended that learners choose their own reading materials (Al-Homoud & Schmitt, 2009:383; Brown, 2009:241), because what is considered enjoyable and easy will vary between them.

The material used for ER is not usually directly studied, so there are no comprehension questions posed (Day & Bamford, 2002:138), and there are no drills to learn the language features in the text (Al-Homoud & Schmitt, 2009:383). As a result, Paran

(2008:467) describes ER as having ‘no focus on the literary qualities of what is being read... and... in its purest form, there is no language learning work either’. Instead, fluency (Hill, 2008:186) and a general understanding of the text (Yamashita, 2004:3) are the immediate outcomes of ER, making it ‘an experience complete in itself’ (Day & Bamford, 2002:138) that limits pedagogic intervention and reduces the teacher’s role to that of facilitator (Hill, 2008:187).

2.2 INTENSIVE READING

On the other hand, Intensive Reading (IR) usually exposes learners to relatively short texts (Hafiz & Tudor, 1989:5) and small amounts of reading (Yamashita, 2004:3), in order to learn language features (Nation, 2004:20), ‘to exemplify specific aspects of the lexical, syntactic or discoursal system of the L2’ (Hafiz & Tudor, 1989:5), and to teach reading skills (Day & Bamford, 2002:136).

With this background, it is usually done in class to allow for a ‘detailed... analysis’ (Bamford, 1984:219) of the texts, which are typically difficult to read unassisted (Nation, 2004:20). To achieve this, IR involves exercises and drills based on the texts, such as answering comprehension questions (Day, 2007:19), completing grammar exercises (*ibid.*, p.20), completing translation exercises (Yamashita, 2004:3), and learning vocabulary (Nation, 2004:20-29).

Hence fluency is sacrificed for a deeper understanding of the texts, making IR a study activity (Waring & Nation, 2004:99) that requires direct instruction. The differences between ER & IR described above and in Section-2.1 are summarised below in Table-1.

Table-1: A Comparison of ER and IR based on the Reviewed Literature in Section-2.1 & Section-2.2

	EXTENSIVE READING	INTENSIVE READING
Length of the reading material	Many pages (i.e. whole books)	Paragraphs, short articles, short extracts
Quantity of the reading material	Large amounts (requiring several hours a week)	Small amounts (requiring a few hours a week)
Purpose of the reading	To give large amounts of L2 input and to improve fluency	To learn specific language features and to learn specific reading skills
Location of the reading	Mainly outside the class	Mainly inside the class
Nature of the reading material	Easy and interesting	Difficult to read unassisted
Selection of the reading material	By the learner	By the teacher
Additional Activities to the reading	Typically none	Comprehension, grammar and vocabulary exercises
Comprehension level of the reading material	The gist or a general level of understanding	Detailed and specific understanding
Teacher's role	Facilitator	Instructor

2.3 ER AND READING

ER is considered an important way to teach reading (Yamashita, 2004:3), given that current research suggests that ER improves reading proficiency (Nakanishi, 2014:6). Others report that ER has been viewed to be the most effective way to improve reading skills (Bamford, 1984:223), while Waring & Nation (2004:106) reported that Krashen went even further with his bold statement that ‘Reading is the only way, the only way we become good readers’. This concept – that ‘students only learn to read by reading; there are no shortcuts’ (Stoller, 2015:158) – is an example of improving proficiency in a language by simply using more of it. In addition, ER has been linked to improvements in reading comprehension and/or reading speed, such as in the studies of Al-Homoud & Schmitt (2009:394), Azmuddin et al. (2014:112), Bell (2001), Elley & Mangubhai (1983:61), and in the studies reported by Al-Homoud & Schmitt (2009:384) and Macalister (2008:248).

All of this is particularly relevant to the context of English for Academic Purposes (EAP), for which good reading skills are essential. It is readily noted that after we learn to read, we then read to learn (Saville-Troike, 1973:396), and so reading has been considered the main way students can independently learn (Schmitt et al., 2011:27)

different scientific, technical, industrial, commercial and academic fields (Elliot, 1962:9). It has been viewed to be ‘one of the most important skills for academic success’ (Schmitt et al., 2011:26), being more important than speaking (Saville-Troike, 1973:395), and even more important than any other language skill (*ibid.*, p.405). The significance of this increases when it has been noted that poor reading skills is usually regarded as the biggest shortcoming in EAP students (King, 1978:38), and hence it has been recommended that ER is incorporated into EAP courses (Benson, 1991:75; Carrell & Carson, 1998:47).

2.4 ER AND OTHER ASPECTS OF LANGUAGE ACQUISITION

The benefits of ER have also been thought to extend beyond the skill of reading, to overall language proficiency as well (Brown, 2009:238). This link is considered intuitive by Hafiz & Tudor (1989:5) and Hill (2008:187), and it has support from research that suggests that ER can improve language learning (Al-Homoud & Schmitt, 2009:383) and language competence (Hafiz & Tudor, 1989:5). One example of this was the long-term study of Elley & Mangubhai (1983), where improvements in language proficiency were described as ‘remarkable’ by Asraf & Ahmad (2003:84). Further reports of other research suggest that ER is one of the best ways to improve language proficiency (*ibid.*, p.83), and even the single most effective way to do so (Brown, 2009:238), although a number of investigations have been criticized in terms of methodology and unclear results (Yamashita, 2004:3).

The basis of ER’s perceived link with improved second language acquisition (SLA) stems from its ability to expose the learner to large quantities of good quality written language input (Elley & Mangubhai, 1983:56). Learners are thought to learn better through good models of language (Cartledge, 1952:96) that reinforce knowledge acquired through direct learning (Hill, 2008:187). This is especially the case in ‘input-poor environments’ (Al-Homoud & Schmitt, 2009:383) where ‘learners have little opportunity to meet and use the L2 outside the classroom’ (Nation, 2003:2), making ER and other forms of input the only non-lesson based opportunities for language

development (Pigada & Schmitt, 2006:2) that can expose learners to vocabulary in their natural contexts (Hafiz & Tudor, 1989:5; Pigada & Schmitt, 2006:2).

In addition, SLA is thought to improve with ER because it has features that mimic first language acquisition. It has been readily noted that many native speakers learn much of their language simply through reading for pleasure (Bamford, 1984:219; Hafiz & Tudor, 1989:4-5), and so it is plausible that second language learners can do the same. It is true that there are important differences between first language and second language learners. Elley & Mangubhai (1983:54-55) summarised these as:

- L1 children have a greater intrinsic motivation to learn, because it is their only language at that stage;
- L1 learners focus more on meaning, i.e. they 'treat language as a tool for achieving some non-linguistic goal, rather than an object to be studied for the purposes of learning a language' (Hampshire & Anoro, 2004:73), while second language learners focus more on form;
- L1 learners receive more language input;
- L2 learners are exposed to a different type of input which is planned, restricted, and largely artificial;
- L1 learners usually get better quality models of language from fluent teachers and authentic literature.

However, it has been proposed that ER of interesting story books can reduce these differences because they 'provide strong intrinsic motivation... and an emphasis on meaning rather than form. When read often, these books increase exposure to the target language... and provide excellent models of written English' (Elley & Mangubhai, 1983:56).

Furthermore, ER seems to comply with a number of other themes in teaching pedagogy that are thought to aid SLA, such as learner-centred teaching, learner autonomy, naturalistic learning, inductive learning, and institutional regime teaching.

Learner-centred teaching is 'based on a now largely unquestioned understanding that learners' *varied* responses to teaching are as important a factor in language learning, if

not more so, than the teaching itself' (Benson, 2004:6) and treats 'language less as an abstract construct of structures and forms and more as a dynamic product of psychological and social life' (*ibid.*). This leads to a focus on both the teacher and the learner, and the interwoven challenges they face, ideally through a process of negotiation that should involve 'joint exploration, consensus building,... renegotiation, equality and mutuality in decision-making and student self-determination' (Martyn, 2000:153) with 'the teacher's aim being to achieve what is best for the students' (*ibid.*, pp.152-153). It has been argued that this can only be achieved by giving learners a greater role in managing their learning, and by giving them more choices, even in terms of methods and scope of study (Littlejohn, 1985:253). The result is that learners become more autonomous, being able to take charge of their own learning, which Cotterall (2000:109) deems essential for all learners and not just the gifted few. This learner autonomy is closely linked to naturalistic learning, which relates to learning outside the formal explicitly taught classroom setting (Benson, 2001:202-203). ER can promote all of this, because it features self-selected reading materials that are mainly read outside the class, and that can 'cater for a wide range of proficiency' (Hill, 2008:187), with each learner choosing according to his/her own level and tastes.

Another important theme in teaching pedagogy that is thought to aid SLA is inductive learning that caters more for holistic learners, and which has arisen out of the shortcomings seen in deductive learning that caters more for analytic learners (Fortune, 1998:68). One example of this is task based learning, where meaningful tasks are used as a vehicle for learners to use L2 in order to complete them and 'only later do they pay attention to language form' (Klapper, 2003:35). Another example is that of data-driven learning, where learners use corpora to investigate language and vocabulary, to 'find answers to their questions... by looking for patterns..., categorizing them and deriving their own hypothesis, rather than relying on a teacher's intuition or research' (Allan, 2009:23). ER shares aspects of these examples of inductive learning, because the learner is exposed to real (written) language in context and then slowly develops an appreciation of its rules and norms. This can be thought of as a consciousness-raising activity that helps learners notice grammatical forms and lexical patterns without the pressure of immediately being required to produce them, and so the common problem of

learners being expected to produce grammatical items before they are ready to do so, is reduced (Fortune, 1998:68).

Teaching according to the practical requirements of the institutional regime (Holliday, 1994:6) is also an important theme in SLA. It has been readily noted that teaching English at tertiary, secondary, and primary institutes outside the English speaking world (termed 'TESEP' by Holliday, 1994:4) is far more challenging than the model situations found in Britain, Australia and North America (BANA), in terms of class sizes and the availability of technology and resources, making some teaching methods propounded by experts from BANA inappropriate (Holliday, 1994:5-6). However, ER is mainly done outside the class, and that which is done in-class can easily be monitored even if class sizes are large, and even if there is little technical equipment.

Hence, ER has been viewed to improve overall SLA both affectively and cognitively (Asraf & Ahmad, 2003:85), and so has been described as being 'pedagogically efficient' (Pigada & Schmitt, 2006:2) because it seems to aid these different aspects of SLA concurrently. In terms of affective aspects of SLA, Brown (2009:238) observed that ER has been thought to improve motivation, with Al-Homoud & Schmitt (2009:398) noting that their ER group 'reported having more positive attitudes towards their learning experience', and with Karlin & Romanko (2010:181), Poulshock (2010:304) and Suk (2016:135) also reporting similar for the ER groups in their more recent studies. In particular, a number of studies have reported that these positive attitudes extend towards second language reading (Al-Homoud & Schmitt, 2009:399; Asraf & Ahmad, 2003:84; Brown, 2009:239; Macalister, 2008:248), with Day (2007:21) asserting that ER is the only way to learn to read and to enjoy foreign language reading.

In terms of cognitive aspects of SLA, it has been thought plausible that ER can improve writing, with reading and writing being input and output respectively, and hence 'two sides of the same coin' that cannot be separated (Elliot, 1962:11). Learning to write through reading has been likened to learning to speak through living amongst (i.e. listening to) native speakers (Bamford, 1984:218), with Hafiz & Tudor (1989:8) explaining that ER gives learners more phrases and better attitudes that leads to a

greater willingness to use those phrases in their writing, and with Waring & Nation (2004:106) reporting Krashen's belief that reading is the only way to develop good writing style. A number of researchers have reported studies that would seem to support the notion that ER improves writing (Asraf & Ahmad, 2003:84; Brown, 2009:238; Macalister, 2008:248), with Hafiz & Tudor (1989:4) and Janopoulos (1986:767) reporting the same in their own studies.

It can be seen from the above that ER has been considered successful in many different contexts and with many different aspects of language acquisition, which has in turn encouraged formal organisations and large-scale collaborations to promote ER. One of these was the Edinburgh Project on Extensive Reading that was set up in 1981 (Yoshizawa, 2014:33), having its origins in large-scale ER programmes that were carried out in East Africa and Malaysia during the 1960s and 1970s (EPER, 1992:4). The Project went on to provide a pivotal role in organising ER programmes in Tanzania, Zanzibar, Hong Kong and the Maldives (*ibid.*, p.5), and it provided additional materials and developed accompanying placement tests and comprehension tests (Yoshizawa, 2014:34). Although the Project ended in 2011, its materials were passed on to ER-Central and the Extensive Reading Foundation, including the database of thousands of GRs that was used by Hill (2008) and Hill & Reid-Thomas (1998) when reviewing GRs, and that are referred to in Section-2.7.1 p.20 and other sections of this thesis. In addition, the Extensive Reading Foundation has extended the work of the original Project in different ways, including storing databases of ER-related materials and research, and including holding mini-conferences and world congresses.

Beyond the Edinburgh Project on Extensive Reading, there have been some other notable large-scale ER projects. The 'Book Flood' project of Elley & Mangubhai (1983) has already been mentioned before in this section, but there were also other such projects in Niue, Fiji, Singapore, Sri Lanka, South Africa, and the Solomon Islands, which appeared to dramatically improve the language skills and enthusiasm of the school children involved (Elley, 2000). A few other projects have been set at higher education settings, such as Robb & Kano's (2013) additive ER programme that

involved over 2500 students at a university in Japan, and the ER programme reported by Tien (2015) that involved over 5000 students at a university in Taiwan.

All of this could indicate that ER has been highly successful in some contexts, and hence supports the argument that ER should have a place in the language learning classroom. Nation (2003:1) suggested that a balanced language course should consist of roughly equal amounts of meaning focused input (i.e. listening and reading), meaning focused output (i.e. speaking and writing), language focused learning through attention to language features, and fluency development by working with known material. ER clearly sits in the first and fourth of these and as such has been recommended to be an integral part of language curricula (Brown, 2009:239; Hill, 1997:58).

2.5 LIMITATIONS WITH ER

Despite the reported benefits of ER, a number of points have been made that highlight its limitations and the research conducted into it. It is important to refer to these, because no matter how plausible and intuitive the benefits of ER may seem, and no matter how convinced a professional may be about them, ‘it would be both unrealistic and shortsighted for a teacher investigating a personally motivating topic to ignore the availability of existing work’ (McDonough & McDonough, 1997:85) in that same topic.

First, there are a number of studies that did not report significant gains in proficiency with ER. For example, although vocabulary gains have been associated with L1 ER (Beglar & Hunt, 2005:8), subsequent studies with L2 learners have yielded only modest gains (Pellicer-Sanchez & Schmitt, 2010:31; Pigada & Schmitt, 2006:1), while Karlin & Romanko (2010:181) found no vocabulary gains after their ER programme. Also, Al-Homoud & Schmitt (2009:393) stated that they found ‘little or no gains in comprehension’ with their ER group, and they mentioned another study that found no improvement in learner attitudes with their ER group (*ibid.*, p.386). In general, a number of investigations that reported beneficial effects for ER have had methodological and curricular limitations (Suk, 2016:20) and a lack of clarity in results (Yamashita, 2004:3).

Second, ER can be viewed to be ineffective with lower-level learners. Day & Bamford (2002:138) describe that a good reader reads faster, reads more, understands better, enjoys reading more, and so reads even faster. However, this cycle takes time to set up in beginners who may not find reading enjoyable, because they do not have enough basic vocabulary to read effectively (Al-Homoud & Schmitt, 2009:387), and because they have not developed the required reading skills that even native speakers take a number of years to acquire (Nation, 1995-6:8). The results of the study of Webb & Chang (2015:651) lend weight to this view, because their lower-level students gained much less vocabulary after the ER programme than their higher-level students.

Third, although it is readily pointed out that reading is an active process (Elliot, 1962:9) that involves ‘an internal dialogue in which hypotheses are formed, predictions made, doubts expressed, uncertainties subsequently clarified, new information grafted on to old, old views modified by new, etc.’ (Williams, 1986:43), it is reported that many students and institutes do not regard ER or silent in-class reading as active learning (Brown, 2009:240; Nation, 2015:143) or even teaching (Macalister, 2008:249). This may stem from teachers and learners being unfamiliar with their role in ER (Brown, 2009:239), and may also result from ER not being part of the Communicative Language Teaching (CLT) that has been popular since the 1970s, and that has focused on speaking and listening at the expense of reading and writing (Hill, 2008:188; Saville-Troike, 1973:395). Although there has been a more critical appraisal of CLT in the past 20 years (see as an example of a useful overview Klapper, 2003), there is still a strong attachment to it, because it is still unlikely that any practitioner would like to be labelled ‘a non-communicative teacher’ (Beale, 2002), and although reading-only lessons can be considered inactive, the same would not usually be said of the conversation-only lessons that are widespread in many places (Dörnyei & Thurrell, 1994:40).

Another cause for this could arise from learners themselves not seeing ER as important in their contexts. Al-Homoud & Schmitt (2009:386) reported a study that showed EAP students had worse attitudes to their ER program than the general English students, perhaps because ER did not seem directly relevant to their specialization, and

Macalister (2008:248-9) noted that ER programmes are particularly lacking in EAP courses, perhaps due to an unspoken belief that it belongs more to schools than tertiary institutes. In addition, many students used to a grammar-translation approach of teaching, may not feel they are getting enough systematic grammar practice (Fortune, 1988:211) from implicit forms of teaching, ER being one of those.

Finally, an important criticism of ER is that it has been deemed time consuming and slow, and hence from this angle, 'inefficient' (Brown, 2009:239). Hence the duration of the ER programme has been considered an important factor for its effectiveness (Mermelstein, 2015:194), but time restraints in real teaching contexts stifle this (Brown, 2009:239), especially in EAP courses (Macalister, 2008:249) and non-BANA settings (Holliday, 1994:4).

All of these reported problems may have contributed to ER being 'almost wholly ignored by textbooks' (Brown, 2009:238), and may have led to its rare inclusion in language curricula (Hill, 2008:189; Macalister, 2008:248), relegating it to a mere 'addition to the language learning programme rather than a central part of it' (Brown, 2009:240), 'a recommended, extra-curricular activity' (Macalister, 2008:248-249), or else simply 'an optional extra' (Hill, 2008:186). More current reviews of research into ER have painted a similar picture of a lack of recent implementation. Day (2015:295) delved into the database of over 500 research articles since 1998 stored in the Extensive Reading Foundation archive, only to find just 44 of them contained both 'extensive reading' in its title and a description of the ER program that was used in its main body. He went on to observe that none of these 44 used all of the 10 principles he co-recommended over 10 year ago in Day & Bamford (2002) and he noted that several discussed arguably the most important recommendation of allowing a free choice of reading, but did not implement it in their programmes (Day, 2015:296). In another review of the same database, Waring (2015:160) noticed that some studies that were labeled extensive reading consisted of reading less than 200 pages during the program, and others consisted of intensive reading of difficult material. Hence the scarcity of recent ER projects was succinctly summarized by Stoller (2015:157) when she observed

that “there remain many L2 settings where there is a total absence of extensive reading, at course and curricular levels”.

2.6 ER AND VOCABULARY ACQUISITION

The relationship between reading and vocabulary acquisition is particularly important. Reading comprehension is thought to depend on a number of factors such as vocabulary knowledge, reading fluency, knowledge of morphology, syntax & discourse, inferencing skills, reading strategies, motivation, and memory (Schmitt et al., 2011:36). In addition, ER requires ease and enjoyment, and this is thought to depend on a text that has a very high proportion of known words, familiar syntax & sentence structure, information that is not too dense, and meaning that is made explicit, especially for beginners (Hill & Reid-Thomas, 1988:45-46). Out of all of these, vocabulary knowledge has been declared to be the most important in affecting text readability (Nation & Beglar, 2007:12), being ‘the most clearly identifiable subcomponent of the ability to read’ and the single best discriminator between good and poor readers (Nation & Coady, 1988:98-101). This need for vocabulary is thought to stem from the observation that understanding written text is not aided by other visible information (such as body language) and audible information (such as intonation) that usually aids understanding spoken language (West, 1964:148-149).

Hence it has been reported that a reader must know at least 95 per cent of the vocabulary in a text to be able to read and comprehend it unassisted, inferring meaning from the context (Schmitt et al., 2001:56), while a number of researchers report that at least 98 per cent of text vocabulary should be known for such easy reading (Clark & Ishida, 2005:226; Day & Bamford, 2002:137).

The importance of vocabulary knowledge is also viewed to extend beyond reading and into other aspects of language proficiency, being ‘directly related to the ability to use English in various ways’ (Schmitt et al., 2001:55). Furthermore, it has been considered ‘essential’ for language learning (Ishii & Schmitt, 2009:5; Schmitt et al., 2001:55), ‘the central component in successful foreign language acquisition’ (Beglar & Hunt, 2005:7),

and it has even been termed ‘a conventional indicator of proficiency’ (Benson, 2001:218). In addition, with vocabulary knowledge being a prerequisite for writing (Li & Schmitt, 2009:85), it has been reported that many ESP learners consider it ‘the most important part of learning a foreign language’ (Schmitt, 1997:3), and that many EAP learners say they need to improve their vocabulary skills (Evans & Green, 2007:14) and they frequently recommend that more vocabulary learning exercises are needed in their courses (Clark & Ishida, 2005:228).

Several studies have been done to ascertain the number of the most frequently used word families that need to be known in order to know 98 per cent of the vocabulary in a text, which as first mentioned above, has been considered the minimal coverage required for unassisted reading (Schmitt et al., 2011:26; Waring & Nation, 2004:98; Waring & Takaki, 2003:135). These studies concluded that the 2000 most frequently used word families i.e. the 2000-word level, is deemed essential for all learners (Wanarom, 2008:43) in order to begin to read the easiest teenage fiction (Nation, 2005b:12); the 5000-word level is needed for authentic unsimplified novels (Hirsh & Nation, 1992:689); and the 8000-9000-word level is needed for a range of authentic texts (Nation & Beglar, 2007:9) such as newspapers (Ishii & Schmitt, 2009:6) and academic texts at university (Nation & Beglar, 2007:12; Schmitt et al., 2011:39). It should be reminded that these refer to ‘word families’, each being defined as a base word with all of its derived and inflected forms that can be understood by a learner without having to learn each form separately (Bauer & Nation, 1993:253; Hirsh & Nation, 1992:692; Kyongho & Nation, 1989:327). Hence the 2000-word level consists of 8418 words (Kyongho & Nation, 1989:327), and the 8000-word level contains 34,660 words (Schmitt et al., 2011:27).

Furthermore, when Schmitt et al. (2011) looked deeper into the original studies that gave rise to the conclusion that for adequate comprehension a reader must know 95 per cent to 98 per cent of the words of a text, they focused on the criteria used to define ‘adequate comprehension’. They reported (*ibid.*, pp.27-28) that in one of these the comprehension test pass mark was 55 per cent, in a later study it was 12/14 for the multiple choice section & 70/124 for the written recall section, while in their own study

(*ibid.*, p.32) the comprehension tests consisted of 14 multiple choice questions and a gap-fill task with 16 spaces. Some of these tests could easily be described as brief, and their criteria could easily be described as modest. The upshot of all of this is that the number of words needed for adequate comprehension is likely to be if anything an underestimate, and so learners are required to learn a very large number of words for unassisted reading (*ibid.*, p.27).

It has been readily observed that it is impossible to directly teach this vast amount of vocabulary in class (Meara, 2005:4; Pellicer-Sanchez & Schmitt, 2010:32; Pigada & Schmitt, 2006:2), especially if the vocabulary is needed for unsimplified texts (Clarke & Nation, 1980:217) or academic study (Macalister, 2008:248). As a result, ER has been proposed to aid this. Waring & Takaki (2003:130) observed that it ‘is received wisdom that people learn most of their vocabulary from reading’, with a number of researchers supporting this by reporting previous studies that seem to show that ER enhances vocabulary acquisition (Al-Homoud & Schmitt, 2009:384; Brown, 2009:238-239; Macalister, 2008:248). This particularly applies to L1 learners (Pellicer-Sanchez & Schmitt, 2010:33), but is thought to apply to L2 learners as well (Meara, 2005:4; Pigada & Schmitt, 2006:1), with the findings of recent studies by Chun et al. (2012:125), Poulshock (2010:304) and Webb & Chang (2015:651) supporting this view. As a result, ER programmes have been recommended as ‘the single most effective way’ (Bamford, 1984:223) to improve vocabulary, and ‘an essential part’ (Nation, 1995-6:9) of learning vocabulary, and as such, Nation (2005a:6) lists ER as one of his ten best ideas for teaching vocabulary.

The term ‘vocabulary acquisition’ is used with ER, because the vocabulary gained during it is thought to occur incidentally (Li & Schmitt, 2009:87; Macalister, 2008:248) where the words are met in context (Nation, 1983:17) and the learner is focused on the story instead of individual vocabulary items (Nation, 2005b:10; Waring & Takaki, 2003:150). In this situation, dictionaries are not used (Saragi et al., 1978:76), and any attention to vocabulary does not interfere much with the flow of reading (Nation & Wang Ming-tzu, 1999:360). Through this, vocabulary is thought to be *acquired*, feeding into ‘the non-conscious memory system of the brain’ (Hill, 2008:187), which differs

sharply from the process of *learning*, which is direct (*ibid.*), explicit (Li & Schmitt, 2009:87) and involves a conscious study of rules and conventions (Hafiz & Tudor, 1989:4).

This process has been observed to be slow (Pigada & Schmitt, 2006:1), leading to only a small quantity of acquired words (Brown, 2009:239; Pigada & Schmitt, 2006:3; Waring & Takaki, 2003:131), with one estimate being a gain of 1000 words from reading 1 million words annually (Al-Homoud & Schmitt, 2009:395), and another being ‘between 150 to 300 words per year’ (Waring & Nation, 2004:107), based on 1-2 hours of in-class reading per week. This slow pace is needed because learners need several exposures to a word before they can acquire even a basic short term knowledge of it (Nation, 2005b:11), with different studies recommending from at least 10 (Wanarom, 2008:43) to over 30 repeated exposures (Waring, & Takaki, 2003:151). Hence, although ‘there is no absolute amount required for reading to be called ‘extensive’’ (Yamashita, 2015:174), a large quantity of reading has been advised in order to achieve vocabulary acquisition through ER (Pellicer-Sanchez & Schmitt, 2010:32), with Nation (2005b:10) advising at least 25 simplified books or 6 unsimplified novels every year, and with Waring & Takaki (2003:155) advising an even greater quantity of 1 book every week.

All of this shows the intertwined mutually reciprocating nature of reading and vocabulary where each is needed to develop the other, such that ‘research leaves us in little doubt about the importance of vocabulary knowledge for reading, and the value of reading as a means of increasing vocabulary’ (Nation & Coady, 1988:108).

2.7 LIMITATIONS IN PREVIOUS STUDIES ON ER & VOCABULARY

It is important at this stage to look deeper into the studies that have been conducted to gauge the link between ER and vocabulary, and to note their limitations in scope, methodology and context. This process, where the research carried out by others is reviewed and synthesised, may be termed as ‘secondary’ research, but is an essential prerequisite in order to appropriately place and plan any new ‘primary’ research (Nunan, 1992:8).

Arguably the most basic limitation in the previous research on ER and vocabulary acquisition is the lack of recent studies relating to the two, which is itself an extension of the situation that relatively few ER programmes in general have been recently carried out, as detailed at the end of Section-2.5 p.13. Hence in a recent article, Nation (2015:139) recommended ER for vocabulary learning based on studies that were all pre-2009, while for this project the researcher found just of handful of relevant studies on ER and vocabulary acquisition that were post-2009 in the database of over 500 research articles since 1998 that are stored in the Extensive Reading Foundation archive (ERF, 2016). Furthermore, one of these recent studies (Azmuddin et al., 2014:112) did not investigate vocabulary acquisition directly, but instead used reading comprehension tests to relate to it indirectly, while another (Poulshock, 2010:315) involved participants being specifically informed that they would be tested on the words found in the books they were reading, potentially putting into question whether the study was investigating incidental acquisition at all. Other recent studies have already been mentioned in earlier sections of the Literature Review, and more about them and other relevant research is referred to below in order to analyze the existing research into vocabulary acquisition through ER.

2.7.1 Limitations in Studies using Graded Readers

Many of these studies have focused on ER using Graded Readers (GRs), which are books that have been specifically ‘written for learners of English using limited lexis and syntax, the former determined by frequency and usefulness and the latter by simplicity... with each stage presenting a more demanding reading task, not only in language but also in length and format’ (Hill, 2008:185). These are often ‘simplified’ versions of novels, but the process of adapting original works involves more than just simplification, and sometimes requires a complete rewrite (West, 1964:146).

Broughton (1962:199-204) explained that in order to tackle the ‘linguistic immaturity’ (i.e. poor vocabulary) of the reader, difficult words are edited out, explained with pictures or a glossary, or replaced with easier words. In order to address the reader’s

‘psychological immaturity’, the story is shortened, chapter headings are made more exciting, spelling is modernized, footnotes and illustrations are used to explain old or culture-specific matters, and unsuitable ideas or their details are censored. To cater for the ‘literary immaturity’ of the reader (who may have little experience of reading), introductions are added to explain the quality of the original story. Furthermore, in the same way course books are usually staged in chapters to help students to learn gradually, these books are also graded into levels of difficulty to provide ‘a ladder up which learners can climb to unsimplified texts’ (Hill, 1997:58), by exposing them gradually to increasingly more complex language forms, and thus providing the effective support and scaffolding for their learning that is considered important for SLA (Hammond, 2006:271), and indeed for learning any subject (Scanlon, 1942:422).

GRs have been found to have ‘an acceptable balance of accessibility and authenticity’ (Allan, 2009:23) because they contain appropriate quantities of lexical chunks, which are considered fundamental to acquire natural and fluent language (*ibid.*), and which are also considered fundamental to write academic texts in EAP (Li & Schmitt, 2009:86). Hence GRs have been used to combine both teaching and enjoyment (Pearson, 1968:245), and they have been considered a primary tool for ER (Hill, 2008:186) because their controlled vocabulary ensures a high repetition of key words, which is considered important for their acquisition (Kyongho & Nation, 1989:333; Pellicer-Sanchez & Schmitt, 2010:35). In addition to GRs being recommended to promote ER (Hill, 2008:198; Williams, 1986:44), they have also been considered the best way to do so (Bamford, 1984:218), and even essential to encourage ER in all except advanced learners (Day & Bamford, 2002:137; Waring & Nation, 2004:99). As a result, GRs have been used in a number of ER studies such as Asraf & Ahmad (2003:90), Hafiz & Tudor (1989:10) and Macalister (2008:250). Similarly, GRs have been recommended for increasing vocabulary (Saragi et al., 1978:73) and for consolidating already known words (Waring & Takaki, 2003:154), and there have been a number of studies into vocabulary acquisition via reading GRs (Al-Homoud & Schmitt, 2009:388; Pellicer-Sanchez & Schmitt, 2010:34; Poulshock, 2010:304; Wanarom, 2008:43; Waring & Takaki, 2003:130; Webb & Chang, 2015:651).

However, it should be remembered that for ER, there should be a large quantity of interesting texts (Nation, 2005b:10-11; Pellicer-Sanchez & Schmitt, 2010:32) that offer repeated exposure to vocabulary (Nation, 2005b:11; Pigada, & Schmitt, 2006:19), especially for beginners. GRs are only a means for this, and in some cases they may not be suitable. Claridge (2012:106) investigated the production rationales of some major publishers of GRs and found a lack of attention to texts that cater for the lowest level of learners, with Pellicer-Sanchez & Schmitt (2010:47) reporting that one study found that many words in GRs were not repeated more than their recommendation of 10 times. It has further been reported that vocabulary in the 3000-word levels and above, appear too infrequently in GRs ‘for reliable learning to occur’ (*ibid.*, p.35), and so readers who rely on GRs will encounter gaps and difficulties when moving up to unsimplified texts (Nation & Wang Ming-tzu, 1999:355; Reid-Thomas & Hill, 1993:252).

Also, GRs may not be easy or interesting to learners. It is readily noted that there can be no one methodology that applies to all times, situations and learners (Klapper, 2003:40), and this observation can be easily made for learning materials as well, because learners are diverse and they differ more in their capacity to learn a second language than their first language (Benson, 2004:20). Hence learners may not find all simplified texts interesting (Ronnqvist & Sell, 1994:126), and sometimes they may find them hard to understand (Elley & Mangubhai, 1983:55), showing that only the learner can judge what is easy and interesting. Learners are not the same and ‘even learners with similar backgrounds vary in terms of the psychological predispositions and learning experiences’ (Benson, 2004:5) and other important dimensions of diversity such as motivation, affect, age, and strategy use (Benson, 2004:20). Teachers can’t expect learners to conform to one approach (Littlejohn, 1985:255), and should avoid sweeping generalizations about their cultures (Littlewood, 2001:21), and instead keep their learners’ needs in mind when making pedagogical decisions (Rounds, 1992:790). This may well result in an observation that many learners are actually not concerned with English literature (Pearson, 1968:245), which is the basis of many western published GRs.

It could be argued that GRs have been especially designed to appeal to most learners, but if it is readily accepted that textbooks can never be perfect for every learner (Nakatsuhara, 2004:1) even though they are usually meticulously designed to cater for students in a particular context or market, then surely GRs will also share this inability to cater for the needs of many learners. Learner choice is an essential aspect of ER (Day, 2007:20; Hafiz & Tudor, 1989:9), and it should always be respected even if that choice is not a GR (Ronnqvist & Sell, 1994:129), because in ER, learners are reading for themselves (Day & Bamford, 2002:137; Williams, 1986:42), and not for teachers, researchers or publishers. Some researchers and practitioners have found learners give positive reviews for GRs, but it should always be remembered that the ‘power relations’ (BAAL, 2010:2.3) between teachers and students could force students to report what they assume may please their teachers, and not what they actually think, and when book reports are written in L2, the students may also not be articulate enough to express their real views clearly (Reid-Thomas & Hill, 1993:265).

In order to address some of these limitations, it has been recommended to provide a large selection of GRs during the ER programme (Hafiz & Tudor, 1989:11, Nation, 1995-6:9), and so for example, Al-Homoud & Schmitt (2009:388) provided 150 GRs for their ER group although it was not clear how many different titles there were, and Pigada & Schmitt (2006:8) provided their participant with a selection of 17 GRs. However, it should always be remembered that ‘it is difficult, if not impossible, to achieve any suitably wide representation of English literatures on a language programme’ (Paran, 2008:488), especially in the case of GRs, which have consistently to be found to be limited in genre. Earlier surveys of GRs (Hill, 1997:57-62; Hill & Reid-Thomas, 1988:47-49; Reid-Thomas & Hill, 1993:252) found that they were predominantly of the fiction thriller genre, based on classics, aimed at teenagers, and featured male protagonists. Although humour and fact-files have become more common, more recent surveys of GRs (Hill, 2008:203) and studies of their vocabulary (Allan, 2009:30-31) have found that fiction still dominates.

In addition, most GRs have also been found to have western contexts and settings (Hill, 1997:62; Hill, 2008:194; Hill & Reid-Thomas, 1988:48), which may not appeal to non-

western learners. Stemming from a need to be learner-centred (Benson, 2004:5-9), it is important that teaching methods, and by extension their materials, are context-specific (Bax, 2003:278; McCabe, 2005:4; Schmitt, 1997:5) if learning is to be optimized, with cultural sensitivity being a key part of that (Bax, 2003:282; McCabe, 2005:6). This can be easily overlooked because, even though ‘we now accept that what may be effective in one classroom with one group of students may not be with another’ (Freeman & Johnson, 1998:402), most teacher training courses give little attention to the different contexts graduating teachers go into (Bax, 2003:281-282; Littlewood, 2000:31). In addition, stakeholders in these contexts and their cultures are often not consulted (Holliday, 1994:3), and so we still see methods that are considered established practice in BANA settings (*ibid.*, p.8) being unsuccessfully transferred to non-BANA settings.

Cultural sensitivity is important because language is used to construct, understand and express thoughts about the world (Bloomer et al., 2005:180-181), and all of this is based on culture. This is even more the case now, with English having emerged as a lingua franca that has in many places expanded at the expense of the local L1, potentially making learners consider their own L1 to be inferior (Nation, 2003:7), and by extension, their own culture as well (Gray, 2000:274-275). This situation also exists in specific fields such as science, for which English is the dominant language (Tardy, 2004:247), making scientists who are non-native English speakers (NNS) disadvantaged when attempting to contribute knowledge, resulting in what has been described as a form of academic imperialism (*ibid.*, p.252). Hence BANA teaching methods, and their associated materials, that are thought to be based on sound pedagogy, could actually be based more on an ideology of ‘cultural chauvinism’ (Holliday, 2007:360), and hence used to preach the western culture and to correct the local one (*ibid.*, p.365), with evidence of this being found particularly in the content of course books (Gray, 2000:274-275). This is sometimes unintended because a number of native speaker (NS) professionals overlook the influence of ideology in their professional lives, even though they are acutely aware of its strong influence in other aspects of their lives, such as in advertising, the media, and politics (Holliday, 2007:364).

The ideology of ‘native-speakerism’, which has been accused of promoting NS teachers and their cultures over others in order to preserve their privileged status (Holliday, 2006:385), has been cited as an important reason for NS teachers to consider their methods and materials from the BANA countries to be superior to those of NNS teachers (Bax, 2003:279), because native-speakerism has led such teachers to fall into ‘culturism’, whereby they reduce foreign people to definitions that those teachers themselves have constructed, but have no existence of their own (Holliday, 2002:45). This can lead to stereotyping and the belief that the NNS culture is lesser than the NS culture (Holliday, 2006:385-386). However native-speakerism, and by extension the belief that teaching methods and materials from the BANA countries are applicable in all contexts, weakens when the issue of the ownership of English is discussed (Holliday, 2006:386). Several decades ago, Pearson (1968:244) noted ‘that English nowadays is not our language alone’ because of its role as an international language for education, science and trade. Now, this ‘outer circle’ of countries (Kachru, 1995:234), where English is learned as the most important new language, has expanded so much that ‘the majority use of English is now outside the English-speaking West’ (Holliday, 2005:8) and hence ‘speakers of English as a first language will lose influence’ (Davies, 2005:4.1) and should no longer expect learners to follow their norms and choices.

With this background, the western context in which many GRs are set (Hill, 2008:194), may innocently seem acceptable to NS publishers and teachers, but may not be so to NNS readers, who may respond negatively to the western cultural setting, just like they could to other affective factors such as subject matter, genre, geographical setting and protagonist gender (Reid-Thomas & Hill, 1993:252). Hence, King (1978:42) noted that ‘materials accepted as adult in Anglo-Saxondom... may seem juvenile or barbarous to those of a different source culture’, while content such as sex, violence, politics, alcohol and drugs, which is acceptable for teenage novels in the West, may be seen as insensitive, offensive and corrupt in other cultures (Hill & Reid-Thomas, 1988:49). Similarly, Hill (2008:192) noted that it is ‘unrealistic, even arrogant, to expect students from non-European cultures to know about the mores and history of western society’, while Saville-Troike (1973:400) gave a number of examples where NNS readers misunderstand culture-specific references because of their own cultural experiences,

noting that this kind of interference ‘frequently goes unnoted and uncorrected by either student or teacher’.

2.7.2 Limitations in Studies using Unsimplified Material

A possible way to address the shortcomings of GRs is to offer an open selection of materials during the ER course, including unsimplified material. This leads to individualized reading which enables learners to read according to their own differing interests and comprehension levels (King, 1978:43). There is a huge variety of unsimplified materials, not just books, but also ‘magazines, newspapers, fiction, non-fiction, texts that inform, texts that entertain, general, specialized, light, [and] serious’ (Day & Bamford, 2002:137). There are also books that are specifically aimed at learners such as teenage books with language and content that is ‘relevant to the life experience, thoughts, emotions, and dreams of young people’ (Ronnqvist & Sell, 1994:126), and books that have L1 translation glossaries, which have been found to be beneficial for text comprehension and vocabulary learning (Laufer, 2005:4).

Hence, it has been reported that unsimplified material can be especially interesting and motivating (Pellicer-Sanchez & Schmitt, 2010:37), complying with ER’s key need to use enjoyable texts (Day & Bamford, 2002:136; Hill, 2008:194) in order to sustain a large amount of reading (Pellicer-Sanchez & Schmitt, 2010:32), because what is read is much less important than the enjoyment derived (Hafiz & Tudor, 1989:9; Williams, 1986:44). If the material is enjoyable, the ‘absorbing content’ (Elley & Mangubhai, 1983:66) can help to overcome the difficulties experienced with unsimplified materials, because learners are human beings and not machines, and as such they have loves and interests that motivate them to persevere and read on (Paran, 2008:469). If the unsimplified material remains difficult to fully comprehend – as could be the case for lower-level learners – it should be noted that ER does not require full comprehension (Day & Bamford, 2002:138; Pritchard & Nasr, 2004), but instead only the gist (Hill, 2008:186) or a general understanding (Yamashita, 2004:3) is required, and so readers have been found to simply skip unknown words if they are not critical to text comprehension (Nation & Coady, 1988:99).

It could be argued that vocabulary acquisition in particular can be better achieved with GRs than with unsimplified texts, especially for lower-level learners. This is based on the premise that GRs are designed to repeat key words (as critiqued above in Section-2.7.1 p.20) and that learners need many repeated exposures to words to acquire them incidentally through reading (Pigada & Schmitt, 2006:19). However, as detailed before in Section-2.7.1 p.20, this is not always the case, with Claridge (2012:106) finding in some major publishers of GRs a lack of attention to texts for the lowest level of learners, and with Pellicer-Sanchez & Schmitt (2010:35,47) reporting that in some GRs the words are not repeated as much as recommended. Furthermore, the quality of attention when meeting a word has been considered more significant in aiding acquisition than the quantity of meetings (Nation, 2015:136), and this can apply to lower-level learners as well, meaning that if they chose exciting and interesting unsimplified material, the resulting extra attention can assist in vocabulary acquisition, even if the words are not repeated so often.

The idea that SLA is mainly pushed by what learners pay attention to is considered 'common sense' by many (Schmidt, 2012:27), with some asserting that noticing is essential for converting input into intake, where noticing requires both attention and awareness (Inzumi, 2002:542). Those who hold this view recognise the possibility and effectiveness of incidental vocabulary learning through reading (Schmidt, 2012:30), but argue that even during this process learners still pay attention, and with more attention comes more incidental learning (Schmidt, 1990:129) and more retention of what has already been learned (Shaw et al., 2010:116). This is particularly relevant given that purely naturalistic learning can be very time consuming (Sharwood Smith, 1981:160), as discussed earlier near the end of Section-2.5 p13, when noting the long duration ER programmes ideally require.

Enhancing attention to vocabulary has often been achieved by enhancing the formatting of the text (Inzumi, 2002:543), even though this is an external measure that is thought to be less effective than techniques that stimulate more internal processes in the learner, such as those stimulated when using the vocabulary in production activities (Inzumi,

2002:566) that are typical in task-based learning (Robinson, 1995:284). Repeating vocabulary is also an attention-enhancing technique (as is aimed for in GRs – see Section-2.7.1 p.20 above), but this is a quantitative technique that is considered less significant than qualitative techniques that require “deeper and more elaborate processing” (Inzumi, 2002:569). Hence interesting and enjoyable books, which are easier to source in unsimplified formats (as argued above at the start of this section), can provide a greater motivation to read that enhances incidental vocabulary acquisition (Shaw et al., 2010:124) by raising the attention of the reader (Schmidt, 2012:40) – including lower-level readers – in a qualitative way that stimulates more internal processes, even when the words are not repeated so often. All of this can lend weight to the argument that unsimplified material can be effective in promoting incidental vocabulary acquisition in even lower-level learners, as long as they find it interesting enough.

A free selection of reading, including unsimplified material, encourages learner autonomy, whereby ‘the learner's perspective is assigned a privileged position’ (Benson, 2001:223), which is important (Cotterall, 2000:109) because it is thought to be a key way to motivate learners (Beglar & Hunt, 2005:9; Dörnyei & Csizer, 1998:215; Pigada & Schmitt, 2006:2). It has been recommended to give learners ownership (Cotterall, 2000:117) and choices (Littlejohn, 1985:253) for how they learn, and ER with a free choice of materials can encourage this, moving books away ‘from *shaping* interests to *catering* for interests’ (Pearson, 1968:243), and aiding learners to read independently, because ‘as teachers of reading our professional objective is to make ourselves redundant’ (Williams, 1986:45).

Theorists have disagreed about how easy the reading material should be for effective ER. Some have insisted it should be within the reader’s ability (Brown, 2009:241; Day, 2007:20; Pigada & Schmitt, 2006:2) i.e. a level easier than the reader’s current ability, while Yamashita (2004:15) described Krashen’s Input Hypothesis as regarding the level above the reader’s ability to provide the condition for acquisition. It could be difficult to provide a selection of books that accommodates both views, and this problem is not necessarily solved even if GRs (with their clearly labelled levels) are used, because it

has been found in one study (Wanarom, 2008:61) that the vocabulary actually used in GRs does not fit closely with the word lists on which they are supposed to be based, and that the levels in series from different publishers do not correspond either. Again, a solution could be to simply allow learners to read books (including unsimplified books) according to the level of difficulty that they choose, especially when considering Al-Homoud & Schmitt (2009:397) found no difference in improved reading speeds or vocabulary acquisition between a group of learners that read books at a comfortable lower level, and another group that read at a more difficult higher level.

It could be construed that by encouraging a free choice of reading, the teacher has no role to play in an ER programme consisting of unsimplified books. However, the teacher can indeed play the role of an informed decision-maker, especially in a context where an eclectic mix of teaching methods are encouraged (Beale, 2002), because teachers are stakeholders that actively participate in the classroom, and as such, possess a great deal of knowledge about their students and their social context (Holliday, 1994:9). Hence the teacher can nurture processes in order to create conditions that are optimal for learning (Beale, 2002) by scaffolding instruction to offer guidance without controlling the decisions learners make (Cotterall, 2000:116), and by opening up a process of negotiation through joint exploration and discussion (Martyn, 2000:153). This could serve better the individual needs and preferences of the learners, whom we cannot expect to follow the same approach to learning (Littlejohn, 1985:255), and whom, even when from similar backgrounds, may differ widely in terms of psychological predispositions, learning experiences, motivation, affect, age, and strategy use (Benson, 2004:5,20).

Hence, in the context of providing a free choice of reading materials during an ER programme, the teacher can still give advice based on experience or research, such as encouraging reading at least two books every month (Nation, 2005a:6), encouraging story books over information-based books (Hafiz & Tudor, 1989:10; Reid-Thomas & Hill, 1993:252), and even giving personal opinions about particular books they feel strongly about (Ronnqvist & Sell, 1994:129). However, with all this, it should be remembered that ER has a particular feature of allowing learners to choose their own

books (Day, 2007:20; Waring & Nation, 2004:105), because they are reading for themselves and not for the teacher (Day & Bamford, 2002:137). Hence the role of the teacher is still active, but the learners' final choice should always be respected (Ronnqvist & Sell, 1994:129), even if that final choice happens to be unsimplified books.

With this background, there would seem to be a case for encouraging a free choice of reading materials (including unsimplified material) during an ER programme. Some have gone further to actively encouraged unsimplified material in order to cover the mid-frequency vocabulary that is under-represented in GRs (Pellicer-Sanchez & Schmitt, 2010:32&35). Others have found that unsimplified material in the form of running newspaper stories can provide the required level of repeated exposure to vocabulary that is needed for acquisition while reading (Kyongho & Nation, 1989:332), which may also by inference apply to other forms of serialised writing. Despite this, only a few of the surveyed L2 ER studies used unsimplified material (Macalister, 2008:251) (Janopoulos, 1986:764), and Pellicer-Sanchez & Schmitt (2010:46) reported that there was little research 'involving unsimplified texts of any kind'. In addition, only a few of the surveyed studies used unsimplified material when investigating vocabulary acquisition through reading (Pellicer-Sanchez & Schmitt, 2010:31) (Kyongho & Nation, 1989:323), and Pellicer-Sanchez & Schmitt (2010:33) reported that there was little such research using authentic novels in particular.

2.7.3 Limitations in Purely Experimental Studies placed in Non-Teaching or Non-ER Contexts

There has been a sustained call for research to take place in real teaching contexts (Holliday, 1994:8-9), and this can also apply to ER research, because if an argument is to be made that it is an effective learning strategy, its research should be placed in the real classroom environment in which it will be employed (Al-Homoud & Schmitt, 2009:386). This firstly arises from the recommendation that ER should be used as just one part of a language learning curriculum (Brown, 2009:240; Hill, 1997:58), and hence it would make sense that research into it is in the context of a complete language

learning course. In addition, learning and cognitive development are culturally and socially based, involving more than just the learner and the teacher (Hammond, 2006:270-271), and so classrooms are the natural place for research into these, because such settings allow different social, anthropologic, psychological and communicative interactions to be displayed and observed (Holliday, 2002:42). Such real teaching contexts should also allow for qualitative research to be conducted, which is important for social sciences like language acquisition, which consists of real human situations, experiences and behaviours that construct inherently subjective realities (Burns, 1999:22). Despite this, Al-Homoud & Schmitt (2009:383) reported that few studies on ER have taken place in real classroom environments.

Furthermore, many previous studies of vocabulary acquisition through reading were not set in true ER conditions. Firstly, a number of studies did not provide a choice of reading material for their participants, and the material chosen was very short and was given a short time to read. For example, Pellicer-Sanchez & Schmitt (2010:37-39) selected just one novel to be read within 1 month, while Saragi et al. (1978:73-74) also chose one novel, to be read within 3 days. Al-Homoud & Schmitt (2009:385) and Pigada & Schmitt (2006:4) reported previous studies with similarly short chosen texts, while Pellicer-Sanchez & Schmitt (2010:35) reported previous studies with durations of reading that ranged from just 10 to 30 minutes.

There are also other limitations associated with some purely experimental studies. It has been recommended that the books in an ER programme should have attractive covers and be displayed nicely to encourage their reading (Bamford, 1984:220). This was done in the study of Elley & Mangubhai (1983:59), and Al-Homoud & Schmitt (2009:388) went further by using a specially made mobile book case to display the books. However, Hafiz & Tudor (1989:9) noted that in such a situation, the participants are well aware that they are on a special programme, and so Elley & Mangubhai (1983:65) mentioned the possibility that the novelty of attractive books may produce a 'Hawthorne effect' (Landsberger, 1958) or 'halo factor' (EPER, 1992:11) that artificially generates positive attitudes and thus may distort the results.

Also, in experimental studies that use widely different treatments between the experimental and control groups, other issues can arise. For example, it has been noted that the benefits of ER does not usually emerge immediately in terms of examination results, but in the study of Al-Homoud & Schmitt (2009:392) the ER group received ‘far fewer and less frequent’ IR exercises, even though such exercises were explicitly described as the basis of the final institutional exams, and yet nothing was mentioned about how they performed in these final exams. Similarly, with it being ‘intuitively plausible’ (Hafiz & Tudor, 1989:5) that ER can improve language acquisition, concerns can be raised about the consequences for control groups during long term projects, like that of Elley & Mangubhai (1983:53) which lasted for almost 2 years, during which the control groups were not exposed to the reported benefits of the experimental group’s ER programme.

In addition, there are the underlying difficulties that exist with any language learning study that uses controls, given that classrooms are not laboratories and so it is almost impossible to ensure that all variables except the one under investigation are held constant between the experimental and control groups (Hyland, 2002:171). One example of this is when different teachers are used, like the study of Elley & Mangubhai (1983:53) that used 16 teachers, because learners often respond just as much to the personality of a particular teacher as to what that teacher is actually teaching (Day & Bamford, 2002:139-140), sometimes leading to ‘self-fulfilling prophecies’ (Littlejohn, 1985:257), because if a teacher is enthusiastic about a certain method, that will be reflected in the way that method is taught, and so the learners are more likely to receive it positively and hence yield beneficial results (Littlejohn, 1985:257).

2.7.4 Limitations in Studies that did not consider Other Sources of English Exposure

It is more than plausible that a significant amount of learning takes place outside the classroom context, being termed ‘naturalistic learning’ by Benson (2001:202-203). This kind of learning can be a cause for a large component of vocabulary acquisition, but it is difficult to observe directly, and so questionnaires and interviews can be used to gauge

it indirectly (Benson, 2001:201). However, no attempt to account for this extra source of language exposure was made in many of the surveyed studies, even those set in rich L2 environments such as Hafiz & Tudor (1989:9) and the studies reported by Pigada & Schmitt (2006:5). Also, a more fundamental point is that although some studies seem to link ER with vocabulary acquisition, few have been able to assert that ER is the cause of that acquisition, because it is still not clear whether this stemmed from increased L2 exposure, or from the fact that the extra exposure was in the form of ER (Al-Homoud & Schmitt, 2009:386). Hence the results from such studies can only usually be stated in terms of correlation and not causation, leading Pigada & Schmitt (2006:2) to conclude that there is still little definitive evidence that ER in itself leads to vocabulary acquisition.

In order to eliminate the possibility of learning target vocabulary from sources other than the reading material, some studies have used texts that contained non-English words that could only be learned from the text and no other source, and so after the reading, the participants would be tested about their knowledge of these words only. Hence Saragi et al. (1978:73) used a book that contained a number of Russian slang words, while other studies used texts with made-up ‘nonsense words’ or ‘nonwords’ as reported by Pellicer-Sanchez & Schmitt (2010:46) and Schmitt et al. (2011:31) respectively, while Waring & Takaki (2003:136) used a set of specially developed ‘substitute words’ that tried as much as possible to follow the normal spelling patterns found in English. However, all of these are unnatural situations that could distort the results, because the non-English words could actually be more obvious and hence be noticed more (Pellicer-Sanchez & Schmitt, 2010:45), or else artificially more confusing and hence harder to learn (Schmitt et al., 2011:40; Waring & Takaki, 2003:152-153).

2.7.5 Limited Consideration of Partial Vocabulary Acquisition

Bitchener et al. (2005:203) reported that SLA research has consistently found that learners acquire linguistic features gradually, and so they sometimes use them correctly and other times they do not, even when the linguistic environment is the same. During this process the learner develops fluency by making best use of what is already known

through tasks that usually involve no new language items and that deal with largely familiar content (Nation, 1995-6:10). Vocabulary acquisition has also been considered to be a gradual process (Li & Schmitt, 2009:97), and not an all-or-nothing process (Pigada & Schmitt, 2006:5; Schmitt, 1998:283), and ER can assist this process by consolidating and deepening the knowledge of partially known words (Brown, 2009:239). Sensitive testing of this partial vocabulary acquisition has been recommended to track its slow and gradual process (Pellicer-Sanchez & Schmitt, 2010:38; Pigada & Schmitt, 2006:2), with the development of such testing being viewed as 'essential' by Schmitt et al. (2001:79).

Despite this, there has been little research into partial vocabulary acquisition (Schmitt, 1999), and few studies have used measurement instruments that were sensitive to this (Pellicer-Sanchez & Schmitt, 2010:35; Pigada & Schmitt, 2006:3), with this absence being thought to be a reason for why some studies have found only small vocabulary gains associated with reading (Nation, 1995-6:7; Pellicer-Sanchez & Schmitt, 2010:33). A more detailed discussion of this can be found in Section-3.4.2 p.75 when describing the methodology adopted in this study to gauge partial vocabulary acquisition.

2.7.6 Limited Research of Lower-Level Male Saudi University Students

A desirable feature of any research is external validity, by which the research design is such that results can be generalised to a wider population beyond the subjects under investigation (Nunan, 1992:14-17). However, in studies relating to ER and vocabulary, the participants may differ widely in terms of psychological predispositions, learning experiences, motivation, affect, age, and strategy use, even when from similar backgrounds (Benson, 2004:5,20), often making external validity difficult to achieve for many contexts.

For example, results from studies with participants who are socially and psychologically similar to those of L2 may not be relevant to learners who are not, because an acculturation model of acquisition would premise that acquisition would be different in the two cases due to the different degrees of social and psychological acculturation

(Benson, 2004:10). Also, ER involves a freer role for both the teacher and the learner that may not be normal in many contexts, especially in places where the text book takes the centre stage of teaching and even takes the place of the curriculum (Brown, 2009:240), or in places where the teacher dominates and is expected to control all aspects of a lesson (Littlejohn, 1985:256). Hence the results of studies with learners from these contexts will again be difficult to apply to learners from a different context, and vice versa. In addition, the very act of reading may be an enjoyable cultural activity for some, but strange and boring for others. Even for many people in the literate West, television has replaced reading as a major leisure occupation (Hill, 2008:188), and the internet has now arguably eclipsed both of them. Hence ER studies with learners who enjoy reading may be far from relevant to the context of learners who do not like reading.

When reviewing the literature, almost all the surveyed studies (such as Elley & Mangubhai, 1983:53; Kyongho & Nation, 1989:323; Pellicer-Sanchez & Schmitt, 2010:36-37; Pigada & Schmitt, 2006:8; Poulshock, 2010:304; Saragi et al., 1978:74; Schmitt et al., 2001:77; Suk, 2016:94; Waring & Takaki, 2003:134) and the studies reported by others such as those reported by Al-Homoud & Schmitt (2009:386) and Asraf & Ahmad (2003:84) dealt with participants who differed greatly from those in this study - in terms of age, nationality, country of residence, country of study, mother tongue, gender, motivation and language ability. Only one surveyed study had similar participants (Al-Homoud & Schmitt, 2009:383) - weak male Saudi students on a university preparatory programme – although the students were at a very different institute to the one in this study, and were studying different majors.

2.8 RESEARCH QUESTIONS

Given the above background, this study focuses on the following research questions:

- Research Question 1 (RQ1) What is the impact of an extensive reading (ER) programme of a relatively short duration, on the vocabulary acquisition of male Saudi university students?
- Research Question 2 (RQ2) What characteristics of the students, in terms of the nature of their reading, their other English learning, and their culture, may also be associated with any observed vocabulary acquisition?

These questions are relevant and ‘intrinsically interesting’ (McDonough & McDonough, 1997:84) because if ER of unsimplified material can be linked to improved vocabulary acquisition in such difficult but natural conditions, then there could be reason to believe that students need not be restricted to GRs, and the case for ER’s wider implementation could be strengthened. Other factors such as the particular traits of their reading, their academic background and their cultural background may also be relevant, and if so, this could have broader applications when developing techniques to enhance vocabulary acquisition, especially in the context of teaching English in Saudi Arabia, which has increased dramatically now that the medium of instruction for all higher secular education is English.

2.8.1 Definitions of Key Concepts in the Research Questions

For both RQ1 & RQ2, the participants of the study were lower-level, poorly motivated male Saudi students who were unacquainted with reading for pleasure even their own language. They were typically 21-22 years old, studying for a bachelor degree in technical majors, and they came from all over the country, including small towns in the poorer regions. A detailed description of the participants is given in Section-3.1.1 p.42

and many of the Results sections in Chapter 4 give further details that vividly describe their social, economic and educational backgrounds.

For RQ1, this study's ER programme was a relatively short 3 month duration, allowed for a free choice of unsimplified and simplified reading materials, and was incorporated into the existing English language preparatory course. In particular, it used what Waring (2015:161) termed 'classical ER' that aimed to fulfil Day & Bamford's (2002:136-139) ten principles for successful ER programmes. This differed from the other kinds of reading that have been associated with ER, such as class reading that entails students reading the same teacher-chosen text (Waring, 2015:161) or reading for pleasure that entails reading material that is enjoyable but not necessarily easy to understand (Beglar et al., 2012:697-698). A detailed discussion of the ER used in this study and how it compares and contrasts the reading used in other studies is given in the middle of Section-3.2 p.54, while a complete account of its rationale and implementation is given in Section-3.2.1 p.57 to Section-3.2.5 p.66. During the ER programme, vocabulary acquisition (as opposed to vocabulary learning – see Section-3.4 p.72 for further details in the context of this project) was investigated using pretests and posttests for the responsive knowledge of spelling & meaning, and the productive knowledge of meaning. There are many aspects to vocabulary knowledge and so the reasoning for selecting these is given in detail in Section-3.4.1 p.73. These different aspects of vocabulary knowledge also served to investigate partial vocabulary acquisition, which as mentioned earlier in Section-2.7.5 p.33 has been considered essential (Schmitt et al., 2001:79) given that linguistic features in general (Bitchener et al., 2005:203) and vocabulary knowledge in particular (Li & Schmitt, 2009:97) are gained gradually. Partial acquisition was also gauged by using vocabulary from different word frequency levels and by asking participants to record their levels of certainty with each response. An in-depth exposition of this is given in Section-3.4.2 p.75.

For RQ2, many aspects of the students' reading were investigated, including the reading material's language, type and topic, as well as each reading sitting's duration, frequency, ease, enjoyment and importance, all of which is detailed in Section-3.6 p.103 and Section-4.4 p.151. Other aspects of the students' English learning were also

investigated, and these included the English they had formally studied in other educational institutes, as well as English they had learned informally during holidays, work placements, and while watching videos and films, as detailed in Section-3.7 p.111 and Section-4.5 p.156. Furthermore, supplementary questioning and in-depth interviews were used to explore aspects of the students' culture that may also be of relevance, as detailed in Section-3.8 p.117, Section-3.9 p.118, Section-4.6 p.162 and Section-4.7 p.164. All of these served to provide a vivid picture of the students of this particular context.

2.8.2 Distinctive Aspects of this Study

With the above background, this study possessed the following distinctive characteristics, attempting to address the limitations that were previously described in Section-2.7 p.19 of the earlier surveyed studies:

1. A relatively short (three-month) ER programme.
2. An ER programme that allowed a free choice of reading material including unsimplified material, and not just GRs.
3. A natural learning and teaching context. This characteristic is important for the sustainability of research into language learning in general (Kluge, 1997:4-6), especially when it has been observed that such research in real classroom environments has been lacking (Holliday, 1994:8-9). In terms of ER, an argument for its effectiveness can only be strong if it is researched in the real classroom environments in which it is to be employed (Al-Homoud & Schmitt, 2009:386).
4. An ER programme that is incorporated into a normal unaltered course of study, which is a common situation for teachers who wish to implement ER, given that in many places ER is not formally included into the language curriculum (Hill, 2008:189; Macalister, 2008:248).

5. A study of partial vocabulary acquisition, noting that vocabulary is not gained in a dichotomous all-or-nothing way (Schmitt, 1998:283), and noting that few studies have been sensitive to this (Pellicer-Sanchez & Schmitt, 2010:35).

6. A study of other factors that may be associated with vocabulary acquisition, including those relating to other aspects of English language learning, noting that few studies have addressed this.

7. A study of poorly motivated, young adult male Saudi university students, who read little even in L1, making the task of implementing an ER programme very challenging.

CHAPTER-3

METHODOLOGY

Research can be considered a systematic process of inquiry (Richards, 2003:9) that requires the collection, analysis and interpretation of relevant data to attempt to answer questions (Nunan, 1992:3). Both internal reliability, which addresses how easily an independent researcher can reanalyse the data to come to the same conclusion (Nunan, 1992:17), and externally reliability, which addresses how easily the whole study can be replicated by others (Burns, 1999:23) are important, but these can only be demonstrated if enough details are given for the chosen methodology in terms of the research contexts, populations, procedures, analyses, and basis for interpretations (Chappelle & Duff, 2003:158-159). This detailing of methodology is also essential to gauge the research's internal validity, which addresses if the research design was such that the outcomes were a result of the specific treatment applied, and to gauge its external validity, which addresses if the research design was such that results could be generalised to a wider population (Nunan, 1992:17). Furthermore, it is recommended that a defensible case can be made for the appropriateness of the particular procedures chosen from the alternatives available (Richards, 2003:4). Hence in this chapter, an overview of the context and methodology of this study is firstly given in Section-3.1 p.42. This is then followed by detailed accounts of the procedures and rationale in Section-3.2 p.54 to Section 3.10 p.123.

3.1 METHODOLOGY OVERVIEW

3.1.1 Development, Context and Participants

The development of this study consisted of three main stages that had different groups of participants, as summarized in Table-2 below:

Table-2: A Summary of the Stages of this Study

STAGE	DESCRIPTION	PARTICIPANTS	ROLE
Stage-1	Initial Pilot	Expert focus group (n=7) of English Language teachers	Piloting the initial versions of the vocabulary tests
		Student focus group (n=4)	Piloting the initial data collection instruments & ER Programme
Stage-2	Main Pilot	Large pilot cohort of students (n=52)	Piloting the revised data collection instruments & ER Programme
Stage-3	Main Study	Experimental cohort (n=49) and Control cohort (n=36)	Using the finalised data collection instruments & ER Programme

The importance of the pilot stages cannot be overstated. Piloting of procedures and data collection instruments is essential to identify and mitigate problems related to clarity, difficulty, administration and processing, with Dörnyei (2003:64) quoting recommendations to abandon studies if the resources for piloting are not available.

The Stage-1 Initial Pilot consisted of two focus groups. In general, focus groups need not be exactly representative of the main study participants (Lederman, 1990:117), but should be a mix of specialist and non-specialist participants who are motivated to give their detailed feedback (Dörnyei, 2003:66). Hence, the first group was an Expert Focus Group of 7 English Teachers, all of whom had master degrees and so were accustomed to research processes. They comprised of 2 Syrian Arabic speakers (fluent also in English), 2 American English speakers, 1 British English speaker, 1 Argentinian Spanish speaker (fluent also in English, German & French), and 1 South African Afrikaans speaker (fluent also in English). Their role was to pilot the initial version of the vocabulary tests, and their broad mix of nationalities, mother tongues and other spoken languages served to provide a rich and varied spectrum of opinions. The second

group was a Student Focus Group of 4 Saudi Arabic speakers. Their role was to pilot the ER programme and all the data collection instruments, as incorporated into their existing preparatory course. Their involvement served to provide views from a student's perspective.

The Stage-2 Main Pilot consisted of 52 Saudi Arabic speaking students, who were from the same social and educational background as the eventual Stage-3 Main Study students. Their role was to pilot the ER programme and all the data collection instruments that had been modified according to the feedback received in Stage-1. This stage enabled all the testing instruments to be checked for statistical reliability, and it also enabled a full trial of the ER programme, all with students that were similar to those of the eventual Stage-3 Main Study.

The Stage-3 Main Study consisted of 85 Saudi Arabic speaking students, 49 of whom were in the ER experimental cohort, and 36 were in the control cohort. These used the finalized versions of the ER programme and the data collection instruments that had been very slightly modified in light of the findings of Stage-2. These numbers (i.e. n-experimental=49, n-control=36, n-total=85) compared favourably with those in other surveyed studies that investigated vocabulary gain through reading, as summarized below in Table-3:

Table-3: Number of Participants in this Study and other Surveyed Studies on Reading and Vocabulary

STUDY	n-Experimental	n-Control
Pigada & Schmitt (2006:8)	1	0
Studies reported by Pellicer-Sanchez & Schmitt (2010:36)	1 for each	0 for each
Pellicer-Sanchez & Schmitt (2010:36)	20	0
Saragi et al. (1978:74)	20	0
Al-Homoud & Schmitt (2009:391)	47	23
THIS STUDY	49	36

Sampling and class allocation was achieved through the institution's English Language placement test, the results of which were used by the institution to allocate the students into classes that had an average and spread of results that were as similar as possible, as detailed later in Section-3.3 p.69.

All of the students in the Stage-3 Main Study (and the Stage-2 Main Pilot) were male, typically 21-22 years old and studying for a bachelor degree in technical majors at the same higher college in Riyadh, the capital of Saudi Arabia. Less than 10 per cent of them were from the capital, and many came from small towns. Most were from the poorer levels of society, with very few being educated in private schools. In addition, very few had parents with jobs that were deemed powerful or influential in the country, such as company directors, security force officers, or high level civil servants.

A number of other aspects of their background indicated that implementing the ER programme would be very challenging. For example, it was significant that they engaged in little reading even in their own L1. Yamashita (2004:2) reported the assertion of some theorists that poor L1 reading attitudes can lead to poor L2 reading attitudes, with her own study (*ibid.*, p.13) giving some experimental support for this, potentially showing the learners in this study may also find L2 reading strange and difficult. Also, their L1 is written in a different script to that of English, which has been thought to make reading more difficult (Elliot, 1962:12), and Arab students have also been thought to have particular difficulty in understanding idioms (O'Sullivan, 1993:173) because many of those in English are specific to a culture that is very different to theirs.

These students also came from a background where textbooks have taken a central place in their learning, with it being given an almost unquestioned authority in providing definitive knowledge (as noted by Harwood, 2005:151, in other countries as well), and their prior learning had been teacher-centred, with deductive methods such as grammar-translation being common place in their English classrooms. Hence it was expected that they may become frustrated with, for example, ER seemingly replacing their preference for explicit grammar teaching (as noted by Fortune, 1988:211 to happen in other countries as well), and that they may view ER sceptically because it is not textbook-based or teacher-centric or used for deductive learning. Hence it was feared that this may lead to less effort being put into the reading and its associated incidental learning (Pigada & Schmitt, 2006:6), and so it was envisaged that a gradual approach would be needed when implementing the ER programme, as is recommended when introducing

any new activity that promotes learner centredness in a context that is usually teacher-centric (Littlejohn, 1985:256).

These students have also displayed a more ‘instrumental orientation’ (as termed by Dörnyei & Csizer, 1998:204) to their learning, usually showing acute concern for course features that are directly related to exam results, but apathy to aspects that are thought to be unrelated. Hence it was anticipated that it may be a struggle to encourage the appropriate enthusiasm for an ER programme that they knew would not be tested for grammar or even comprehension (as also observed by Fitz Gerald, 1947:359, in other contexts).

Overall, these students, in studying for a bachelor degree that had been formulated and taught by western teachers, were being bombarded with methods, styles and priorities that were very different to what they had been used to in their prior educational culture. ER was just one of these new components, and like with EAP in the context of students who are new to the western learning culture, it was expected that acculturation training would need the same attention as language proficiency development (as noted by Errey et al., 2004:5, in other settings).

3.1.2 Overview of Procedure during the Main Study

During the Main Study, the following overall procedure was implemented. The adopted methodology was based on considerations of those used in previous studies, and was based on the findings and lessons learned from the Initial Pilot and the Main Pilot studies, all the time ensuring a careful and systematic approach was being followed. All of this is described below and then detailed further in Section-3.2 p.54 to Section-3.10 p.129 of this chapter.

By way of background, the institutional exams at the higher college that was the setting of this study were based heavily on the course books, which required about 80 per cent of the lesson time in order to cover their content. During the remaining contact time, teachers were expected to supplement with appropriate work. A part of this remaining

time was used to incorporate the ER programme with the experimental cohort. This was very much an innovation and a departure from the norm, given that reading for pleasure was atypical, and allocation of it during a formal course of study was even more so, making its implementation very challenging. During the control cohort's remaining lesson time, other English learning activities based on the normal course were provided. This situation meant that both groups had equal quantities of English learning contact time, but the experimental cohort had some of that in the form of ER. This mitigated the problem highlighted by Al-Homoud & Schmitt (2009:386) that in some previous studies, the experimental ER groups simply had more English exposure than the non-ER control groups who had no extra English teaching in any form, and so unsurprisingly recorded higher gains in proficiency. More details of the procedure for the experimental and control groups are given in Section-3.2 p.54.

In order to address both RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), this study's ER programme possessed a number of features. First, it was incorporated into the normal course of study for the experimental group, responding to the call of some for more research on individual learners that is set in natural contexts (Benson, 2004:11). ER is ideal for this, being viewed as something that can be easily blended into any language learning curriculum regardless of focus (Day, 2007:20), with Hill (1997:58) actively recommending its integration. Another key feature was that the experimental cohort was given an open choice of reading materials, including unsimplified materials. A free choice of reading has been listed as an important principle of ER that allows learners to choose what they find interesting (Day & Bamford, 2002:137) and appropriate to their level of comprehension (King, 1978:43). In addition, the ER programme had in-class Sustained Silent Reading (SSR) sessions, following the recommendation of Day & Bamford (2002:139), and following its use in some studies such as that of Elley & Mangubhai (1983:59). Further details of the ER programme are found in Section-3.2 p.54.

Many data collection instruments were adopted to address research questions RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2

(concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). These are now summarised below in Table-4, with most of the data collection for the Main Study being completed in 2013.

Table-4: Data Collection Instruments during the Main Study

	DESCRIPTION	QUANTITY	RELEVANT RQ	DETAILED IN:
1	Institution English Placement Tests	n=85 scripts	Sampling & Class Allocation for RQ1 & RQ2	Section-3.3 p.69 & Section-4.1 p.134
2	Pretests and Posttests developed by the researcher, to gauge different aspects of Vocabulary Acquisition	n=680 scripts for 4 separate tests, gauging 3 different aspects of partial vocabulary knowledge, and testing 200 words	RQ1	Section-3.4 p.72 & Section-4.2 p.135
3	Institution End-of-Semester English Exams	n=85 scripts	Check effect of ER Programme on Participants' exam results	Section-3.5 p.101 & Section-4.3 p.149
4	Reading Diaries developed by the researcher to record student reading patterns	n=49 diaries, recording an aggregate of approximately 600 weeks of reading	RQ2	Section-3.6 p.103 & Section-4.4 p.151
5	Surveys developed by the researcher to record Other Sources of Exposure to English	n=49 surveys	RQ2	Section-3.7 p.111 & Section-4.5 p.156
6	Supplementary Questioning	hundreds over a period of 2 years	RQ1 & RQ2	Section-3.8 p.117 & Section-4.6 p.162
7	In-Depth Interviews	12 participants (approx. 25% of the Experimental Cohort), for approximately one hour each	RQ2	Section-3.9 p.118 & Section-4.7 p.164
8	A Research Journal containing notes, observations and events	spanning the duration of the project	RQ1 & RQ2	Section-3.10 p.129

The timeline for the procedure used during the Main Study's teaching semester is summarized below in Table-5, showing the intensive nature of the whole programme that was filled with the administration of the ER programme and the data collection instruments.

Table-5: Timeline for the Procedure during the Main Study Semester

Week	EVENTS in the MAIN STUDY	DETAILED IN:
1	Institution English Placement Tests	Section-3.3 p.69
1-2	Pretests of the different tests of Vocabulary knowledge	Section-3.4 p.72
2	First week of the ER Programme and recording of Reading Diaries	Section-3.2 p.54 and Section-3.6 p.103
6	First whole class discussion on reading to date	Section-3.6.4 p.108
9	Mid-semester break	Section-3.2.2 p.57 & Section-3.2.5 p.66
10	Second whole class discussion on reading to date	Section-3.6.4 p.108
11	Survey of Other Sources of Exposure to English Learning	Section-3.7 p.111
12	Whole class discussion about Survey of Other Sources of Exposure to English Learning	Section-3.7.4 p.115
14	Last week of the ER programme and recording of reading diaries	Section-3.2 p.54 and Section-3.6 p.103
14-15	Posttests of the different tests of Vocabulary knowledge	Section-3.4 p.72
17-18	Institution end-of-semester English Course Exams	Section-3.5 p.101

In order to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36), pretests and posttests of vocabulary knowledge were administered to both groups to gauge vocabulary gain. Most of the surveyed studies also used testing, with for example the Vocabulary Levels Test (Schmitt et al., 2001) being used in so many studies that it has been described “as close to a standard test as we have for this purpose” (Al-Homoud & Schmitt, 2009:389-390). Hence both the Vocabulary Levels Test and another widely used test (the Vocabulary Size Test of Nation, 2009) was used so that the results of this study could be compared and contrasted more easily with those of other studies. It should also be noted that the 200 target words used in this project’s vocabulary acquisition tests (see Item-2 above in Table-4) were considerably more than those tested in a number of surveyed studies (see Table-6 below), potentially allowing for a greater depth of investigation.

Table-6: Words Tested in this Study and other Surveyed Studies on Vocabulary Gain through Reading

STUDY	NUMBER OF TARGET WORDS TESTED
Waring & Takaki (2003:134)	n=25
Pellicer-Sanchez & Schmitt (2010:38)	n=34
Al-Homoud & Schmitt (2009:390)	n=90
Saragi et al. (1978:73)	n=90
Pigada & Schmitt (2006:1)	n=133
THIS STUDY	n=200 (see Section-3.4.3 p.76)

Further details of the vocabulary tests and how their design helped to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36) are given in Section-3.4 p.72, and given that the 200 tested words yielded a huge amount of data, an in-depth statistical analysis (that is comprehensively described in Section-4.2 p.135) was chosen to process the results.

However, to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36) a greater variety of data collection instruments were considered necessary. Hence reading diaries were completed to allow for a detailed investigation of the students' reading and how that may have affected their vocabulary gain, while a survey of other sources of English learning was administered to explore how these may also have been relevant. Most of the participants' reading habits and all of their English learning outside the college were impossible to observe directly, making diaries and surveys excellent instruments to collect such data. In addition, large amounts of quantitative data arose from these, justifying a statistical analysis of their results. An exhaustive description of how the methodology adopted for these two data collection instruments aligned with RQ2 is given in Section-3.6 p.103, Section-4.4 p.151, Section-3.7 p.111 & Section-4.5 p.156.

Furthermore, in-depth interviews were also used to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). These were chosen because the data that arose was qualitative, and so this complemented the more quantitative data that arose from the diaries and surveys. Also, the in-depth interviews gave the participants an opportunity to express their own ideas in their own words, making any arising conclusions more personal and pertinent to the exact context at hand. The methodology of the interviews is further explained in Section-3.9 p.118, and the analysis of the resultant data is fully detailed in Section-4.7 p.164, with both sections explaining the rationale of the procedures adopted in light of RQ2.

In addition throughout the entire project, hundreds of supplementary questions were posed to the participants, giving insights that moulded the entire study and hence addressed both RQ1 (concerning the impact of the ER programme on vocabulary

acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). These were presented in an informal manner, which allowed the participants to be relaxed and candid in their responses, and they were immediate in nature, which allowed responses to be accurately recorded. In some situations the questions followed a set procedure, for example when checking entries in the other data collection instruments, while at other times they were spontaneous, for example when clarifying an unexpected or interesting response. The implementation of this questioning is further expanded upon in Section-3.8 p.117. During the pilot stages, the responses to these questions led to important amendments to the other data collection instruments that were eventually used in the Main Study, and the procedure for analysing the data collection instruments was also moulded by the feedback obtained through these questions, all of which is detailed more in Section-4.6 p.162.

Overall, this methodology follows what Benson (2001:196) reports to be frequently used by researchers, in that they follow up quantitative data with qualitative data, given that

‘bringing together (triangulating) multiple perspectives, methods, and sources of information... adds texture, depth, and multiple insights to an analysis and can enhance the validity or credibility of the results ...’

(Chappelle & Duff, 2003:165).

Examples of this in the field of vocabulary testing include Schmitt et al. (2001:57) who used interviews to validate the responses and scores of their participants in the Vocabulary Levels Test they were developing, while in the field of extensive reading Macalister (2008:251-254) used quantitative questionnaires with qualitative interviews to investigate (amongst other things) the participants’ attitudes to the ER programme.

3.1.3 Research Design

This study’s research questions RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36) called for different methods of investigation, and so the research design had to accommodate for these. Nunan (1992:4-6) described research

design to involve different combinations of 3 parameters: design (experimental or exploratory), nature of data (quantitative or qualitative) and analysis (statistical or interpretive). He (Nunan, 1992:5-7) also described research design in terms of the two parameters of interventionism (the extent to which a researcher intervenes in the environment) and selectiveness (the extent to which a researcher pre-specifies the phenomena to be investigated). With this in mind, different parts of the study had different research designs.

Research question RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36) was addressed by implementing an ER programme with one cohort of students, and comparing their vocabulary gains with that of another cohort that had no ER programme (see Table-4 p.47 & Table-5 p.48 above). Hence this aspect of the study, because of the pretests, posttests and treatment (T1–treatment–T2 as described by Al-Homoud & Schmitt, 2009:391) can be considered purely ‘analytical-nomological’ (Nunan, 1992:6) with its experimental design, quantitative data and statistical analysis, and because of its high level of interventionism and selectiveness (Nunan, 1992:7).

However, for research question RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36) a different approach was adopted. The use of the Reading Diaries and Surveys of Exposure to English Learning (see Table-4 p.47 & Table-5 p.48 above) best fitted a mixed research design (Nunan, 1992:4), because they were exploratory in nature, i.e. non-experimental in design, but the data yielded was quantitative and its analysis was statistical (see Section-4.4 p.151 & Section-4.5 p.156), while the Supplementary Questioning and In-Depth Interviews (see Table-4 p.47 above) can be considered a purely ‘exploratory-interpretive’ (*ibid.*, p.6) because of their non-experimental design, qualitative data and interpretive analysis, and because of their lack of interventionism and selectiveness (*ibid.*, p.7) (see Section-4.6 p.162 & Section-4.7 p.164).

Statistical correlation analyses using multiple regression and structural equation modelling were also considered to address research question RQ2 (concerning the

impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). However the researcher wanted in-depth information about the participants of this particular context, and was not explicitly seeking to gain a simplified generalisation that could be offered by such correlation analyses. Hence the Reading Diaries, Surveys of Exposure to English Learning, Supplementary Questioning and In-Depth Interviews (see Table-4 p.47 above) were favoured, in order to achieve the desired detailed rich description of the students in this context, which in many cases used the students' own voices, and in this way the very human and social nature of this particular part of the study was preserved as much as possible.

This study also shared characteristics of more than one broad research tradition. For example, it shared some key aspects of the action research tradition described by Burns (1999:24-30), because it was conducted in a naturally occurring setting, it addressed questions of real practical and theoretical interest to many educational practitioners, it placed practicing teachers in the research process, and it aimed to bring about change and improvement to day-to-day teaching practices through the data collected, and not only through the views of non-practitioner theorists. At the same time it was not a full action research project, because the study had broad implications that went far beyond the limited professional and educational contexts of its teachers, students and college. Also, in order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), a full scale immersed ethnography was not essential, but nonetheless there were still some characteristics of ethnography, such as Chappelle & Duff's (2003:174) prerequisite that a researcher has spent a long time residing and interacting with the study setting (which in the case of the researcher was many years), and Hyland's (2002:196) criteria of being contextual (being set in the natural learning environment) and longitudinal (taking place over months). Some of Hyland's (*ibid.*) other criteria for ethnography were also fulfilled to a significant extent, such as that the study should be emic (privileging the perspectives and choices of participants) and unobtrusive (avoiding manipulating the phenomena as much as possible). This study even has some limited aspects of grounded theory when investigating RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), because the data collected from the reading diaries, surveys and interviews was used inductively

to allow ideas and theories to emerge organically (Richards, 2003:16-17) concerning other factors that could be associated with enhanced vocabulary acquisition.

Hence in summary, some of the data collected was more quantitative in nature, while other data was more qualitative, taking the view that both are complementary and equally important (Burns, 1999:24), and that the distinction between the two is not always clearly defined (Nunan, 1992:3; Richards, 2003:11): the quantitative data was predominantly used to give insights into the characteristics of the participants (i.e. the *what*, *how much* and *how many*), while the qualitative data was chiefly used to ascertain the reasons for those characteristics (i.e. the *why*).

3.2 METHODOLOGY FOR THE ER PROGRAMME

This section describes in detail the implementation of the ER programme during the Main Study, and how it was based on recommendations given in the literature, and then further developed during the Initial Pilot and Main Pilot Studies (see Section-3.1.1 p.42). The main principles followed were the ten given by Day & Bamford (2002:136-139), namely that the ER programme should consist of:

1. Material that the students find easy to read
2. A variety of different reading materials to cater for different interests
3. A free choice of reading materials
4. Opportunities to read lots of reading material
5. A focus on reading for pleasure and information
6. An aim to make reading its own reward and an experience in itself
7. A focus on increasing reading speed
8. Opportunities for silent reading
9. Teacher guidance for good reading practice
10. Teacher models for good reading practice

Day (2015:295) termed any programme that implemented all these principles as ‘pure’ ER, with Asraf & Ahmad (2003:87) and Macalister (2008:250) aiming to implement all of them in their ER programmes, and with other researchers recommending at least some of them. For example, Nation (2001:3-4) echoed the need to encourage large quantities of reading, and Waring & Nation (2004:105) further noted that large quantities of easy student-chosen reading material would maximize the effectiveness of the ER programme in enhancing vocabulary acquisition. In addition, Littlejohn (1985:260) highlighted the importance of teachers’ attitudes, in order to facilitate successful independent learning (of which ER is one form), and Paran (2008:478-480) highlighted the importance of the methods used for teaching and presentation, in order to maintain learners’ interests in the literature put forward to them. Others were more explicit in their insistence that the teachers implementing the ER programme should themselves feel positively about ER (Asraf & Ahmad, 2003:99), and should themselves

be enthusiastic and committed to reading (Bamford, 1984:260). As an additional point, Macalister (2008:254) noted that although the above 10 recommended characteristics are considered solid principles, there still needs to be flexibility in approach when implementing the programme.

It should be pointed out that there has been a great deal of variance in the way ER has been implemented, as recently noted by Macalister (2015:122) and as found by both Day (2015:295) and Waring (2015:160) in current reviews of hundreds of research articles on ER. For example, Day (2015:296) noted that several ‘ER’ studies did not allow a free choice of reading, and Waring (2015:160) observed that some ‘ER’ programmes involved little reading – sometimes less than 200 pages – with other examples being given earlier in the Literature Review (Section-2.7.3 p.30). Such a variety of application may be a source of confusion that leads to differing viewpoints about ER (Macalister, 2015:122) and so there has been a call to define ER more precisely into a common framework that can be referred to by all (Waring, 2015:165).

The previously quoted 10 principles of Day & Bamford (2002:136-139) have served to be a basis to define ER, but recent efforts have been made to further refine the definition of ER into core features without which the term ‘ER’ should not be applied. In this vein, Macalister (2015:126-127) reduced the 10 principles to 7, where the reading is: for pleasure, information & general meaning; fast; individual & silent; of easy material; guided by the teacher; exemplified by the teacher’s own reading; and done as much as possible in a regular time-limited period, noting that Day & Bamford’s 10 principles did not include the length of time for which students should be reading. Waring (2015:161-165) went further to reduce ER into four core elements that the reading should entail: fluent comprehension (for which beginners are allowed time to develop); a high speed; large amounts; and a focus on meaning, while Yamashita (2015:169) went even further to hold that large amounts of reading were ‘the essence of ER’ that can only be achieved when the reading is enjoyed (*ibid.*, p.173).

As a result, reading programmes can be described more precisely with terms like ‘classical ER’ (Waring, 2015:161) for programmes that use all the previously listed 10

principles of Day & Bamford, and ‘modified ER’ or ‘light ER’ (Day, 2015:296-297) if many of them or a few of them respectively are used. However, other terms can be used when core elements of ER are not present, such as class reading that involves a single teacher-chosen text read by the whole class (Waring, 2015:161), or reading for pleasure that involves reading material that is enjoyable but not necessarily easy to understand (Beglar et al., 2012:697-698) in the sense that 98% of the vocabulary may not be known (Strong & Boutorwick, 2012:71).

Hence in this study, the ER programme aimed to use ‘classical ER’, attempting to follow Day & Bamford’s principles, and it also aimed to use Macalister’s (2015:126-127) addition criteria that the reading should be done as much as possible, thus automatically including the core features quoted above from Waring (2015:161-165) and Yamashita (2015:169). An outline of this ER programme is now given below in Table-7, followed by a full detailed description in Section-3.2.1 to Section-3.2.5.

Table-7: Timeline for the ER Programme of the Main Study Experimental Cohort during its Teaching Semester

Week	PROCEDURE for the ER PROGRAMME	DETAILED IN:
2	Introduction of ER and its aim of reading for pleasure Statement that no tests would be made on the reading Teacher guidance given on what is easy and interesting reading material Solicitation of students’ existing reading interests Presentation of a large selection of books available at local book stores Instruction to buy 2 books for the next lesson and bring more when finished	Section-3.2.2 p.57
2-13	Sustained Silent Reading sessions planned for 20 minutes every day Encouragement (but no compulsion) to continue reading outside the class	Section-3.2.3 p.61 & Section-3.2.4 p.65
9	Mid-Semester Break	Section-3.2.2 p.57 & Section-3.2.5 p.66
10	Formal introduction to the institute’s library of GRs	Section-3.2.2 p.57
14	Last week of the ER programme	Section-3.2.5 p.66
17-18	Institution end-of-semester English exams	Section-3.5 p.101

3.2.1 Incorporation of the ER Programme into the Normal Course of Study

In this study, the ER programme that is detailed throughout this Section-3.2 was incorporated into the normal unaltered course for the experimental cohort. It should be remembered that ER is considered an approach that can be blended into any language course (Day, 2007:20), and it has been recommended that ER should be one of many integral parts (Hill, 1997:58) of a course, reflecting theory and research that recommends an eclectic mix of teaching methods (Beale, 2002). Also, incorporating ER into the normal course was a prudent way of helping learners become gradually accustomed to ER's unfamiliar characteristics, with this principle of introducing new ideas concurrently with existing familiar ideas being recommended for text books (Brown, 2009:244), and also being equally advisable for methods of teaching. This context of ER being incorporated into a pre-existing course was challenging, but not atypical of many ER programmes. There are many barriers to ER's formalized and official implementation, such as syllabus demands, time constraints, and possible doubts in the eyes of administration about its benefits (Macalister, 2008:249), and so in many contexts, the teacher is frequently left alone to attempt to integrate it into the curriculum, and to deal with the practicalities involved (Brown, 2009:238).

3.2.2 Open Choice of Reading Materials

The learners in the experimental cohort were given an open choice of reading material, actualizing this essential characteristic of ER that makes it more likely that the material read will be interesting (Day, 2007:20), and following the few examples of surveyed studies that did offer a free choice of reading (Janopoulos, 1986:764; Macalister, 2008:251). It was explained that the material required should be both easy (as stated by Day, 2007:20, to be an essential aspect of ER) and interesting (as recommended by Hill, 2008:194 and Nation, 2005b:11), and that without interesting texts, little sustained reading would be possible (Pellicer-Sanchez & Schmitt, 2010:32; Williams, 1986:42), emphasizing that the reading should be for pleasure and not for study like the other aspects of their curriculum. This was further reiterated by informing the students (as

advised by Day & Bamford, 2002:139) that perfect comprehension was not needed, and that no tests would be given to check comprehension.

Some minimal teacher guidance was then given about how to determine if a book was easy and interesting, noting that teachers usually do have good knowledge of their students and their social contexts (Holliday, 1994:9), and noting that teachers have valued judgements about books that should be shared (Ronnqvist & Sell, 1994:129). In this way, the teacher nurtured the students by using knowledge and experience to provide the optimum conditions for ER, just like a gardener provides the optimum conditions for a plant to grow, while not forcing it (Beale, 2002).

Hence in terms of choosing easy books, the students were advised to look for the following features:

1. Short books with many pictures. These help to address what Broughton (1962:201) termed the 'psychological immaturity' of the reader. A short book is easier to finish, and so is more motivating because of the associated feeling of achievement (Hafiz & Tudor, 1989:10), while pictures help to explain the text, and serve to reduce text density, which further motivates readers to read more. This was exemplified by the results of an investigation reported by Hill (1997:65) that showed short books with many pictures were borrowed more often than long books with few.

2. A clear large font. Small print has been judged to be 'the single most off-putting factor' (Hill, 1997:66), and small spaces between lines have been judged to make 'any page unreadable' (Hill & Reid-Thomas, 1988:45).

3. Introductions, footnotes and glossaries. These help to address what Broughton (1962:202) termed the 'literary immaturity' of the reader, who may have little experience of reading literature. Hence introductions help to describe the background of the story and its literary worth, while footnotes and glossaries help to describe culture-specific terms and difficult words.

4. Known vocabulary. Hill & Reid-Thomas (1988:46) asserted that vocabulary knowledge is the most important requirement of an easy text, because reading speed and enjoyment is reduced when frequent reference to a dictionary is needed. Hirsh & Nation (1992:689-690) added that even knowing 90 per cent of the vocabulary would still result in approximately 1 unknown word for every line, and so 98 per cent or more has been recommended (Day & Bamford, 2002:137). Hence, to impress this upon the students, they were first asked if they thought 90 per cent vocabulary knowledge would be enough for easy reading, to which they always agreed it would be. Then 10 lines were drawn on the board and the students were informed that they represented 10 lines of text, with each line containing about 10 words. Then a red mark was made on each of the lines, representing 1 unknown word per line. This visual representation was used to demonstrate how many times a reader would need to use the dictionary for the unknown words. Then, the students were asked how long they would take to look up a word, even with an electronic dictionary, and then multiply this time by 10 to give the total time looking up unknown words from just 10 lines of reading. This combined visual and mathematical representation completed the demonstration that 90 per cent word knowledge would not be enough for easy reading. Instead, the students were instructed to test-read a few pages of any book they were thinking of reading, and they were told that if there were more than a few unknown words in a given page, then they should not choose it, unless those unknown words were not essential to overall text comprehension, and would not need reference to a dictionary. This provided a simple and precise way for the students to gauge if a potential book was suitable in terms of containing known vocabulary.

5. Repeated vocabulary, as is often found in story books. Nation (2005b:11) reported that vocabulary gained through reading is gradually lost unless it is reinforced with repeated exposures. Hence novels are often recommended (Hirsh & Nation, 1992:689) because their continuous texts provide opportunities for this. Other recommended material includes biographies and travel books (Reid-Thomas & Hill, 1993:252) because their strong narrative element can offer the same opportunities, as can 'running' newspaper stories, defined by Kyongho & Nation (1989:323) as original news stories with their follow-ups reports.

In terms of choosing interesting books, it has been recommended to survey what the students enjoy in L1, in terms of the books they read (Day & Bamford, 2002:137) or the programmes and films they watch (Hafiz & Tudor, 1989:10), and then guide them to similar material in L2. Hence the students of the experimental cohort were first asked about their general reading habits in L1. They readily admitted that did not read much for pleasure, with the exception of daily newspapers. This gave the opportunity to advise them to try reading English newspapers, especially those printed locally and so covering news of local concern which they would be already aware of through their Arabic newspapers. After this, they were asked about the programmes and films they watched, and it was interesting to find that they did watch English films, but with Arabic subtitles. This enabled the teacher to quote the book that was the original source of a particular film (e.g. Sherlock Holmes), or the book that was published following the release of an original film, such as ‘movie story books’ that usually contain an extended summary of the plot, glossy photographs of different scenes, and features about the characters and actors.

However, in this study’s ER programme, canvassing the students’ reading habits was only the beginning of introducing interesting texts. Hence many different books that were available in the local bookshops were presented to and passed around the students, aiming to show them that they should be able to find something interesting, even if their interests were not shared by others. Day & Bamford (2002:139) recommended this introduction to a wide range of reading materials, and gave examples such as ‘books, magazines, newspapers, fiction, non-fiction, texts that inform, texts that entertain, general, specialized, light, serious’ (*ibid.*, p.137). Hence the books that were shown included: story books; picture story books; movie story books; books about science, space, wildlife, animals, cars, and planes; magazines about computers, smart devices, football, and business; comics and annuals; newspapers; graded readers; and novels with full Arabic and English translations on facing pages. Some of the books had content that was clearly aimed at young adults, reflecting the observation that it ‘is well documented in the literature’ that learners react better to literature that is aimed at their own age group (Paran, 2008:488). However, books that could be considered more

suitable for children were also presented, as were other books with content more suited for older adults, simply to give the students as many ideas and options as possible.

In addition, different versions of the same book were on display to show that if the unsimplified novel version was too difficult, they could read the GR version, or the Arabic-English version, or even the picture story version. Also, it was shown that some books were available electronically from the internet, and downloadable onto smart devices. The prices of the books were also given, emphasizing that many of the books were very cheap. Each time the teacher showed a different kind of book, he reiterated that it was only an example of what was found in the local bookshops, and that it was down to the student to choose for himself. It should be noted that the college did have a library of GRs, but it was feared that if this was formally shown at the beginning of the programme, the students would naturally restrict themselves to these easily obtainable books at the expense of the vast variety of other reading material available in the shops. Instead, the GR library was only formally introduced after the mid-term break (see Table-7 above), by which time the students had already bought their own varied selection of books.

The students were then initially instructed to buy two books of their choice and bring them to the class the next day to read, and when they had finished they would have to get more. They were also informed that they did not need to complete a book if they did not enjoy it, and instead they had the complete freedom to change or swap books whenever they wanted, as recommended by Arnold (2008:98) and Day & Bamford (2002:137) for students in any ER programme.

3.2.3 Sustained Silent Reading Sessions

It has been asserted that for ER, the reading should be silent (Day & Bamford, 2002:139) and individual (King, 1978:43), being termed ‘sustained silent reading’ (SSR) by Asraf & Ahmad (2003:86), who also reported other terms used in the literature, such as ‘pleasure reading’ and ‘free voluntary reading’. Williams (1986:46) advised that the primary aim of any reading class should simply be to assist the students

to read, and not to spend the time listening, discussing, or testing comprehension. Even reading aloud has been discouraged, because it is thought to slow reading speeds and inhibit comprehension (Griffin, 1992:784), due to the extra concentration required to pronounce the printed words (Elliot, 1962:15). Hence, in this study's ER programme, in-class SSR sessions were planned every day, noting that in general most students need in-class reading sessions to be encouraged to read more outside class (Asraf & Ahmad, 2003:87; Bamford, 1984:221; De Burgh-Hirabe & Feryok, 2013:90), and noting that a number of other studies have also used them, such as Al-Homoud & Schmitt (2009:392), Asraf & Ahmad (2003:90), Elley & Mangubhai (1983:57), and Macalister (2008:250).

The role of the in-class SSR sessions in starting and encouraging the reading habit, was particularly important in the context of this study's participants, who read little even in L1. Nation (2005b:10) pointed out that reading itself requires considerable knowledge and skill, and even native speakers are unable to gain much vocabulary through reading during the time they take to develop their reading ability (Nation, 1995-6:8). Also, they were unfamiliar with a reading activity in which that they had more control than the teacher. Hence the SSR sessions were an example of an appropriate way to gradually introduce what was to them a new form of learning, as recommended when guiding learners to take responsibility for managing their own learning (Littlejohn, 1985:261).

Asraf & Ahmad (2003:94) observed that the participants in their study needed time to become accustomed to silent reading, and this study's participants were no different. During the Initial Pilot & Main Pilot Studies, it was noted that in the first week of in-class SSR sessions, the students needed to be actively trained to sit silently and concentrate on reading. It was clear that some were trying to purposely break the monotony by asking questions, requesting to go to the toilet, looking through the window, etc. Students were reminded that the reading class involved reading and nothing else, and even requests to borrow a dictionary were not allowed. In such cases, the requesting student was simply advised to carry on reading, even when a word remained unknown. If a student complained that his book was difficult or boring, he was reminded that he was the one who chose it, and if the difficulties and boredom

continue, he would have to replace it by the next lesson. In these first few sessions, the students were told that the aim was to become absorbed in the book, and that this would require at least 10 minutes of complete concentration.

Also, during the Pilot stages, some students were seen to initially make copious notes while reading, and would look up every word in a dictionary. In response to this, dictionaries were never forbidden, but the students were reminded that their comprehension would not be examined, nor would any grammar be tested. Again, this took some time for the students to really believe, because their previous educational experiences were heavily teacher-centric and directly linked to formal summative assessments.

It was clear to the researcher (who was also the observing teacher) that some would pretend they were concentrating, with tell-tale signs of frequent furtive glances towards the teacher to see if they were being watched. Hence gentle reminders were given that the reading was solely for pleasure, and hence independent of the formally assessed course. A sign of success in encouraging concentration during the SSR sessions was to see the students focused on their books and wanting to carry on reading, even when it was announced that the allotted time had finished.

The Initial Pilot & Main Pilot Studies were also very valuable in trialling the logistics of the in-class SSR sessions. The students had 6 English lessons every week (i.e. a lesson every day, except one day that had two), each being 1 hour and 40 minutes long. In-class SSR sessions were first held at different times during the lesson, and it quickly became apparent that holding them at the end of lesson was not ideal, as the students seemed to be tired. Also, they initially lasted 40 minutes, but it became evident that this sometimes restricted the time needed to cover the main course material, and it was also too long for many students to hold their concentration. During the weeks, other durations were tried, and in the end, 20 minutes seemed ideal to allow the students to become absorbed with their reading, without becoming bored, and without using too much lesson time. In addition, 2 sessions were initially held every week, but more were tried out in subsequent weeks, to which the students seem to respond positively. This,

along with the modest significant vocabulary gains found during the Main Pilot (see Section-3.4.7.7 p.100 and Appendix-15) supported the decision to try to increase the number of in-class SSR sessions for the Main Study.

Hence in the final Main Study, the in-class SSR sessions were fixed at 20 minutes duration, held during the middle of the lesson, and every attempt was made to hold them each day, although in the end an average of 3.53 sessions per week was achieved. This was similar to other surveyed studies that implemented an ER programme, as shown below in Table-8:

Table-8: Sustained Silent Reading (SSR) in this Study and other Surveyed ER Studies

STUDY	Duration (min)	Frequency (sessions/week)	Part of the Lesson
THIS STUDY	20	5 (planned); 3.53 (average achieved)	Middle of 100min lesson
Al-Homoud & Schmitt (2009:391-392)	20-25	4	Middle of 50min lesson
Elley & Mangubhai (1983:57)	20-30	5	Not indicated
Macalister (2008:250-251)	20	5	Not indicated
Asraf & Ahmad (2003:90-91)	40	5	During 80min lesson
Hafiz & Tudor (1989:7)	60	5	Whole 60min lesson

It has often been recommended to mix ER with IR activities (Al-Homoud & Schmitt, 2009:392; Williams, 1986:44), including in EAP contexts (Carrell & Carson, 1998:47). Also, Nation (2005b:15) recommended supporting any ER programme with language focused learning and fluency development, while others advised that explicit vocabulary learning activities are needed to accompany ER (Laufer, 2005:3; Pellicer-Sanchez & Schmitt, 2010:43; Pigada & Schmitt, 2006:21). In addition, Nation (2005b:13) recommended follow up activities such as oral book reports, discussion groups and book awards, with Day & Bamford (2002:138) adding further examples including writing about favourite characters, or the best or worst books, or doing a dramatic reading of an exciting part of a novel.

However, in this study's ER programme, minimal extra activities were done, fearing that such activities would be affective, especially in the context of weak learners who

were not accustomed to reading for pleasure. It was hoped that this approach would mitigate the anti-motivational effects of any activity that could be perceived as checking up on students or testing them (King, 1978:42), or making them feel forced to read (Paran, 2008:477), and it was instead hoped that this would promote better learner autonomy (Pigada & Schmitt, 2006:2). Hence, the students were simply left to read, noting that this should be the main aim of any ER programme (Nation, 2001:3; Nation, 2005b:12), as implemented by Elley & Mangubhai (1983:59), Hafiz & Tudor (1989:7) and Macalister (2008:250) in their studies. It should be noted that dictionaries with L1 translations were allowed during the in-class SSR sessions. This has been explicitly recommended by Lott (1960:67) and Laufer (2005:4), especially in the context of lower level learners (Ishii & Schmitt, 2009:10) who are still at the stage of learning the most frequent 2000 word families (Nation, 2003:4).

3.2.4 Encouragement to Read Outside the Class

The students were then encouraged to continue reading after class, as advised by Nation (2001:3) to be essential once the reading programme has been initially set up. They were given examples of places and times they could continue reading, based on their typical routines, such as when waiting for the beginning of a lesson in college, when having coffee in a café in the evening, or when flying back to their home cities during the weekend. They were given confidence by constantly reminding them that reading would get easier, due to its ‘virtuous’ cycle (Day & Bamford, 2002:138) in which when a reader reads more, he understands more, then enjoys more, then reads faster, and so reads even more. The students were also encouraged with realistic targets such as a minimum of 1 book per week (as advised by Day & Bamford, 2002:138), and with the teacher actively displaying a positive and enthusiastic attitude to reading and being a good example by reading during many of the SSR sessions (as advised once again by Day & Bamford, 2002:139-140). Hence, beyond the in-class SSR sessions, the students were encouraged, but never forced to read, hoping this tension-free atmosphere would aid acquisition (Hafiz & Tudor, 1989:4).

3.2.5 Duration of ER Programme

Nation (2001:2) noted that ER programmes should be as long as possible, if proficiency gains are to be seen. In this study, the experimental and control cohorts' semesters consisted of 16 teaching weeks, which was clearly the upper limit duration for any ER programme. However, it was also important that all the course material was explicitly covered, and in particular, the last two weeks were very important in terms of syllabus completion and revision for the final exams. However, during the Main Pilot, only modest significant vocabulary gains were found (see Section-3.4.7.7 p.100 and Appendix-15), and so for the Main Study the ER programme was lengthened to 13.1 weeks, including the 1.4 week mid-semester break, i.e. 11.7 weeks excluding the break. This was comparable to some other ER studies, as shown below in Table-9:

Table-9: ER Programme Duration in this Study and other Surveyed ER Studies

STUDY	Duration of Reading Programme
Saragi et al. (1978:74)	3 days
Pigada & Schmitt (2006:1)	1 month
Al-Homoud & Schmitt (2009:392)	10 weeks excluding a 3 week mid-semester break
Macalister (2008:251)	10 weeks
THIS STUDY	13.1 weeks including a 1.4 week mid-semester break, i.e. 11.7 weeks excluding the break
Hafiz & Tudor (1989:7)	12 weeks
Asraf & Ahmad (2003:99)	4 months
Elley & Mangubhai (1983:53)	20 months

3.2.6 Control Cohort

Both the experimental and control cohorts studied the same course in preparation for the same formal assessments held at the same times throughout the semester. As described in Section-3.1.2 p.45, the course material required about 80 per cent of the lesson time, and during the remainder, teachers were expected to supplement with appropriate work. A part of this was used to incorporate the ER programme with the experimental cohort, while the control cohort engaged in other English learning activities, and so both groups had equal quantities of English learning contact time, but the experimental cohort had some of that in the form of ER, while the control cohort had no explicit ER programme.

In terms of details, the teacher of the control cohort used this extra time he had after explicitly covering all the formally examined course material, to facilitate the students with the following:

- making posters to display language that was presented in the course material;
- giving verbal presentations of the above posters to the rest of the class;
- completing word searches of vocabulary contained in the course materials;
- completing crossword puzzles containing the vocabulary in the course materials;
- generating whole-class discussions based on the contents of the course materials;
- practising conversations based on the contents of the course materials;
- using teacher-generated worksheets to do extensive practice of the grammar related to the present simple, past simple and present continuous tenses;
- using teacher-generated worksheets to explain writing structure at sentence level only;
- providing extensive practise of writing affirmative, negative and interrogative sentences.

Hence in summary, it can be seen from the above that the teacher of the control cohort mainly used the extra time he had to do more work with the vocabulary and grammar presented in the course materials but in different ways and with different resources. As mentioned before above, this was in contrast to the equivalent time used by the experimental cohort, which was instead used for the formalised ER programme.

It should also be pointed out that some of the surveyed ER studies involved many different teachers, such as the large scale study of Elley & Mangubhai (1983:53) that had 16 teachers in the experimental groups and 8 teachers in the control groups (*ibid.*, p.59), which could have made ensuring consistency in teaching approach and attitude extremely difficult. However, in this study, the experimental groups had the same teacher (i.e. the researcher), and the control groups were all taught by one other teacher, with whom a close liaison was possible throughout the study. This was deemed an important characteristic because teachers, like learners, are important participants in a classroom (Holliday, 1994:9) and different teachers' attitudes are as important as learners' attitudes in affecting the outcome of a course (Littlejohn, 1985:257). Hence, by ensuring that the experimental groups had the same teacher and that the control groups

had the same teacher, it was easier to mitigate as much as possible, the potentially large effect of different teaching attitudes and styles upon the different groups, and hence maintain as consistent a methodology as possible.

3.3 METHODOLOGY FOR THE PLACEMENT TEST

Participants from the Main Pilot (n=52) and the Main Study (n=85) (see Section-3.1.1 p.42) completed the institution's English Language placement test during their first lesson, and this was used as a convenient tool for sampling and class allocation, thus being a preliminary stage of the project.

For reasons of confidentiality that were assured when taking permission to carry out this study, a copy of the test cannot be included here. However, it can be described as containing 40 multiple choice grammar and reading comprehension questions, based on the Preliminary English Test (PET), which is 'an intermediate level qualification... [relating to] English language skills for work, study and travel.' (Cambridge English Language Assessment, 2014). The students were given the lesson (1 hour 40 minutes) to complete it, and all were able to do so within this time.

The institution then allocated the students into their classes to ensure as much as possible that the average and spread of these results were similar. The method adopted was that their scores were first ranked, and then the first student (i.e. the one with the highest score) was placed in class A, the second in class B, the third in class A, the fourth in class B, etc. This simple procedure was to ensure a similar distribution of proficiencies in both the Experimental and Control groups, and this was demonstrated with the statistical calculations that are detailed below in this section. It could be raised that the placement test was not necessarily an appropriate indicator of proficiency, and hence not suitable to use for sampling. However during the Main Pilot Study, it was readily observed by the researcher and other teaching colleagues that the results of the placement test did usually match their own informal assessments of the students' proficiency during the initial weeks of the course.

It was not necessary to demonstrate that the scores should be exactly the same for the Experimental and Control cohorts, because RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36) focused on the change in vocabulary knowledge of the participants, as calculated through the difference between

their pretest and posttest results. However, all the participants should have been at least approximately the same level of proficiency, because for reading (and hence for acquiring vocabulary through reading), advanced learners have a significant advantage over lower level learners, who may not have yet developed an enjoyment for reading, nor enough basic vocabulary to read effectively (Al-Homoud & Schmitt, 2009:387), nor other reading skills that even native speakers take years to acquire (Nation, 1995-6:8).

Excel spreadsheets and IBM's SPSS (v.17 and higher) were used to process the data to yield results for both the descriptive and inferential statistics (see Appendix-14). The descriptive statistics (Pallant, 2010:59-63) showed that the data was not normally distributed, and so the non-parametric Mann–Whitney U Test for 2 independent samples was used (*ibid.*, p.227), where $p < 0.05$ indicated significant difference between the groups (*ibid.*, pp. 240-2).

The above-mentioned Mann–Whitney U tests (see Appendix-14) showed:

1. There was no significant difference between the Placement Test results of the Main Pilot cohort (Median=41%, $n=48$) and its total batch (Median=40%, $n=593$), $U=13948$, $z=-0.23$, $p=0.818$.
2. There was no significant difference between the Placement Test results of the Main Pilot cohort (Median=41%, $n=48$) and the Main Study Experimental cohort (Median=45%, $n=45$), $U=1023$, $z=-0.44$, $p=0.658$.
3. There was no significant difference between the Placement Test results of the Main Pilot cohort (Median=41%, $n=48$) and the Main Study Control cohort (Median=33%, $n=32$), $U=597$, $z=-1.68$, $p=0.093$.
4. There was no significant difference between the Placement Test results of the Main Study Experimental cohort (Median=45%, $n=45$) and the Main Study Control cohort (Median=33%, $n=32$), $U=549$, $z=-1.77$, $p=0.077$.

The above Result-1 demonstrated that the participants of the Main Pilot were not significantly different to the rest of their batch, and so they could be taken as a good representation of the students at the college. This gave confidence that the methodology and data collection instruments developed during this stage would be applicable to the eventual Main Study. Result-2 and Result-3 confirmed these assumptions.

The above Result-4 demonstrated that the participants of the Main Study Experimental and Control groups were not significantly different. This gave confidence that it could be valid to compare the two after treatment, and that any differences arising would not be those that existed beforehand.

3.4 METHODOLOGY FOR TESTING VOCABULARY GAIN

In this study, pretests & posttests of vocabulary knowledge were developed and administered to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition – see Section-2.8 p.36). It is important at this stage to define vocabulary acquisition in the context of this study, because there has been a great deal of variance in studies that have investigated vocabulary gain over the years, with examples of this being given immediately below in Section-3.4.1 p.73. This variety could be considered natural given that vocabulary knowledge has many different facets and given that vocabulary can be gained in different ways. However, this can also be a source of confusion that leads to different viewpoints about the effectiveness of ER on vocabulary gain, thus making it important to explain vocabulary gain more precisely so that like-for-like comparisons can be made with other studies.

Hence in the context of this project, the word ‘vocabulary’ (in the term ‘vocabulary gain’) is defined with three facets: the responsive knowledge of spelling, the responsive knowledge of meaning, and the productive knowledge of meaning. Vocabulary knowledge has many different facets, and the full rationale for choosing these for this study is detailed immediately below in Section-3.4.1 p.73.

Furthermore for this project, the word ‘gain’ (in the term ‘vocabulary gain’) is defined with two components. Firstly, it relates to incidental gain, i.e. vocabulary *acquisition*. As first mentioned in the Literature Review (Section-2.6 p.16) vocabulary acquisition occurs when words are met in context (Nation, 1983:17) and the focus is on the story (Nation, 2005b:10; Waring & Takaki, 2003:150), feeding into ‘the non-conscious memory system of the brain’ (Hill, 2008:187). This contrasts with the process of vocabulary *learning*, which is direct (*ibid.*), explicit (Li & Schmitt, 2009:87) and involves a conscious study of rules and conventions (Hafiz & Tudor, 1989:4). The methodology adopted to ensure as much as possible that the words in the pretests were not subjected to any special attention or learning is detailed below at the end of Section-3.4.6 p.89. Secondly, the word ‘gain’ relates to *partial* gain, noting that vocabulary acquisition is gradual (Li & Schmitt, 2009:97) and not an all-or-nothing process (Pigada

& Schmitt, 2006:5), as first expounded in the Literature Review (Section-2.7.5 p.33). This was achieved for the context of this study in three ways: (1) by testing different facets of word knowledge, namely spelling and meaning, as noted above in the previous paragraph; (2) by testing vocabulary from different levels of word frequency; and (3) by seeing the level of certainty participants reported when giving their answers. The detailed justification and procedure for all these three ways is comprehensively described below in Section-3.4.2 p.75.

Hence what follows now is a detailed account of all of this, including the tests themselves, their administration and their rationale, connecting the methodological choices with the research questions, and hence aiming to display what Nunan (1992:10) considered key characteristics of experimental research: that it should be systematic, logical, tangible through the use of real world data, replicable, and reductive in order to establish patterns among the observable phenomena.

3.4.1 Testing Specific Aspects of Word Knowledge

Fluent use of vocabulary requires both breadth and depth of vocabulary knowledge (Ishii & Schmitt, 2009:6-12), with breadth being the number of words recognised, and depth being further knowledge of those words, including inflections, shades of meanings, collocations and restrictions of use (Waring & Nation, 2004:101-102). This has been broken down further with Ishii & Schmitt (2009:6-7) and Schmitt (1998:285) reporting the receptive and productive aspects of 8 categories of word knowledge: spoken form, written form, grammatical behaviour, collocational behaviour, frequency of use, stylistic register constraints, conceptual meanings, and associations with other related words. In addition, Nation (2004:22) provided more detail by giving 9 categories, each with receptive and productive aspects, as given in Table-10 below.

Table-10: Facets of Word Knowledge (Nation, 2004:22)

CATEGORY OF WORD KNOWLEDGE		RECEPTIVE ASPECT	PRODUCTIVE ASPECT
1. FORM	Spoken	What does it sound like?	How is it pronounced?
	Written	What does it look like?	How is it spelled?
	Word parts	What parts are recognized?	What parts express meaning?
2. MEANING	Form & meaning	What does it mean?	What word form can be used to express this meaning?
	Concepts & referents	What is included in the concept?	What items can the concept refer to?
	Associations	What other words are thought of?	What other words could be used instead?
3. USE	Grammatical functions	In what patterns does it occur?	In what patterns must it be used?
	Collocations	What other words occur with this?	What other words must be used with this?
	Constraints	When, where and how often is it met?	When, where and how often is it used?

With this multi-faceted nature, complete knowledge of a word requires much more than just its form and meaning (Schmitt, 1998:283), and as a result, it has been noted that no one test has yet been developed to gauge all these forms of lexical knowledge (Ishii & Schmitt, 2009:7; Schmitt et al., 2001:61), but instead many studies have only been able to focus on one of these forms (Waring & Takaki, 2003:133). For example, Pellicer-Sanchez & Schmitt (2010:34) reported some studies that only investigated spelling, while Pigada & Schmitt (2006:1-2) observed that many other studies have only investigated meaning, such as the studies of Al-Homoud & Schmitt (2009:390), Saragi et al. (1978:73) and Waring & Takaki (2003:130). A few of the surveyed investigations looked into additional aspects of word knowledge, such as grammatical characteristics (Pellicer-Sanchez & Schmitt, 2010:34; Pigada & Schmitt, 2006:2) and word associations (Schmitt, 1998:281), but for most studies, vocabulary gain has typically been gauged through testing knowledge of the form-meaning relationship of the target words (Waring & Nation, 2004:102).

Out of all the different facets of word knowledge, the receptive skills are thought to develop first (Elliot, 1962:14; Saville-Troike, 1973:400; Schmitt, 1998:285), with recognizing the form of the word coming before working out its meaning (Waring & Nation, 2004:103). This has led to the assertion that making the form-meaning

relationship (matching the spelling of a word with its meaning) is the first facet of word knowledge acquired when reading (Waring & Nation, 2004:101), and the results of some studies lend weight to this (Pigada & Schmitt, 2006:19; Schmitt, 1998: 281). This is particularly relevant to the context of this study's participants, because they were lower-level learners who read little even in their own L1. This meant that form and meaning were likely to be more relevant aspects of their vocabulary learning than the other facets of word knowledge, and hence likely to be acquired first.

All of this background served to validate the construct (Nunan, 1992:15-16) that vocabulary acquisition in this study was defined in terms of the scores obtained by testing form and meaning only.

3.4.2 Testing for Partial Vocabulary Acquisition

One of the distinctive aspects of this study (see Section-2.8.2 p.38) was to gauge partial vocabulary acquisition, because vocabulary acquisition has been viewed to be a gradual process (Li & Schmitt, 2009:97; Pigada & Schmitt, 2006:5; Schmitt, 1998:283) that can be enhanced through ER (Brown, 2009:239) (Section-2.7.5 p.33). This has led to a call for research, and vocabulary tests in particular, that are sensitive to this gradual acquisition (Pigada & Schmitt, 2006:2; Schmitt et al., 2001), although few have emerged (Pellicer-Sanchez & Schmitt, 2010:35; Pigada & Schmitt, 2006:3; Schmitt, 1999).

One aspect of gradual learning relates to the progressive acquisition of all the facets of word knowledge, with for example, recognizing the form of the word before being able to understand its meaning (Waring & Nation, 2004:103). One of the few surveyed studies that attempted to address this was that of Pellicer-Sanchez & Schmitt (2010), which investigated knowledge of spelling, word class and meaning, and which also used half marks for word class answers and meaning answers that were incomplete according to the criteria set by the researchers (*ibid.*, pp.38-40).

Another aspect of gradual learning (as described in Section-2.6 p.16) is that more frequently occurring words are more likely to be learned before less frequently occurring words, as was found by Waring & Nation (2004:101). Hence for instance, the 2000 most frequently occurring word families are likely to be learned before others.

Finally, the level of certainty possessed concerning any one facet of word knowledge is an aspect of gradual learning. At one end of the continuum, learners may not know a word at all, but may still feel compelled to guess in a test situation, resulting in an answer that may be correct, but is not a true reflection of their knowledge. This problem has been poorly addressed in many tests, such as the earlier version of a vocabulary level test made by Nation (1983:15) which did not account for any possibility of guessing. In other tests, participants were simply instructed not to guess (Pigada & Schmitt, 2006:10; Schmitt et al., 2001:66), while some later tests did formally account for guessing by asking participants to select a ‘don’t know’ option for each unknown item (Ishii & Schmitt, 2009:11; Pellicer-Sanchez & Schmitt, 2010:38; Schmitt et al., 2011:43). As for the rest of the spectrum, Schmitt (1994:13-14) reported just one test that gauged certainty of each word according detailed descriptors, ranging from level 1 (‘I don’t remember having seen this word before’) to level 5 (‘I can use this word in a sentence’).

Given the above background, the tests in this study were designed to account for partial vocabulary acquisition in three different ways:

1. Testing different facets of word knowledge
2. Testing vocabulary from different levels of word frequency
3. Accounting for different levels of certainty for any given answer

Further details of this are given below in Section-3.4.3, Section-3.4.4 and Section-3.4.5.

3.4.3 Overview of Vocabulary Tests Used

In the same way that there can be no one best method for learning that applies to all contexts (Klapper, 2003:40), there can also be no one best way to test learning (Waring

& Nation, 2004:103), and in terms of testing vocabulary, the kind of test has been found to affect the results obtained (Waring & Nation, 2004:102; Waring & Takaki, 2003:133). Written tests with multiple choice questions (MCQs) are widely used for language research (Dörnyei, 2003:43), and they have been used in many vocabulary acquisition studies like the studies reported by Waring & Takaki (2003:149) and the surveyed studies of Pellicer-Sanchez & Schmitt (2010:38), Waring & Takaki (2003:133) and Saragi et al. (1978:73), while others have used interviews (Pellicer-Sanchez & Schmitt, 2010:39) or dictated tests (Pigada & Schmitt, 2006:10). With this background, it has been explicitly recommended to use different tests (Waring & Takaki, 2003:134).

Hence, in this study there were 4 tests, which are summarised below in Table-11, and further detailed in Section-3.4.4 p.79 & Section-3.4.5 p.86.

Table-11: Summary of the Vocabulary Tests Used in This Study

	Test-1	Test-2	Test-3	Test-4
Layout of Script	Combined with Test-2	Combined with Test-1	Separate Script	Separate Script
Word Knowledge tested	Receptive Written Form i.e. SPELLING	Productive Form-Meaning i.e. OPEN TRANSLATION	Receptive Form-Meaning i.e. MEANING	Receptive Form-Meaning i.e. MEANING
Kind of Test	Multiple Choice	Open Response	Multiple Choice	Multiple Choice
Source of Target Words	Vocabulary Size Test (Nation, 2009)	Vocabulary Size Test (Nation, 2009)	Vocabulary Size Test (Nation, 2009)	Vocabulary Levels Test (Schmitt et al., 2001)
Total Number of Target Words tested	n=50	n=50	n=50	n=150
Word Frequency Level Groupings	1000-5000 & 6000-10000	1000-5000 & 6000-10000	1000-5000 & 6000-10000	2000, 3000, Academic, 5000 & 10000
Target Words per Grouping	25 & 25	25 & 25	25 & 25	30, 30, 30, 30 & 30
Word Class Ratio Noun:Verb:Adjective	26:14:10	26:14:10	26:14:10	3:2:1
Administration	First (combined with Test-2)	First (combined with Test-1)	Second	Third

The 4 tests shared some common features (see Appendix-3 to Appendix-10). Firstly, for all the tests, the instructions used English that was as simple as possible. This was essential, because it has been noted that slight changes in wording can often lead to

misunderstandings and varied responses (Dörnyei, 2003:32-33), and these are even more likely with instructions presented in a foreign language (Benson, 2001:196) to lower-level learners (Chappelle & Duff, 2003:167). Hence the instructions were in short sentences, using language that was free of jargon, abbreviations and other technical terms, as advised by Dörnyei (2003:52-55), and clear examples were also included. All of this was in addition to the verbal explanation and demonstration given to the participants during test administration (see Section-3.4.6 p.89). Furthermore, all instructions had alongside them Arabic translations, which were checked for accuracy and ease of understanding with the Arabic speaking participants of the Initial Pilot Study (see Section-3.1.1 p.42).

Another feature shared by all the 4 tests was the use of Likert scales to measure the level of certainty associated with a response, with certainty levels being one aspect of partial vocabulary acquisition under investigation in this study (see Section-3.4.2 above). The term ‘certainty’ is inherently subjective, and so it was quantified and operationalised (Burns, 1999:21) by using Likert scales because of their simplicity, versatility, and reliability (Dörnyei, 2003:36). An even number of options was chosen to prevent participants taking the middle category that perhaps avoids making a thoughtful choice (Dörnyei, 2003:37) and instead of the normal left to right arrangement, the numbers were arranged from right to left, following the reading direction used in the students’ L1 of Arabic. Initially a 6 point scale was used (as recommended by Dörnyei, 2003:38), with ‘0’ representing an answer that was a blind guess, and integer intervals up to ‘5’ that represented an answer the participant was absolutely sure was correct. However, following feedback during the Initial Pilot Study, this was changed in the revised versions of the tests to a 4 point scale, from ‘0’ to ‘3’, with ‘3’ representing absolute certainty (see Section-3.4.4.1 p.82 for further details).

Finally, all the 4 tests had words grouped into different levels of word frequency, with this also being one of the three ways that this study gauged partial vocabulary acquisition (see Section-3.4.2 p.75 above).

What follows now is a detailed description of the contents and development of each of these tests, and then an overall account is given about how they were administered.

3.4.4 Overview of Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning)

In this study, Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) used words from Nation's Vocabulary Size Test (2009), a multiple choice meaning test whose basis was described by Nation & Beglar (2007:9-13). This original test contained vocabulary from the 1st 1000 to the 14th 1000 most frequent word families of English (i.e. from 1000-word level to 14000-word level), although it was not intended that a learner need be tested for all these levels (*ibid.*, p.11). Word families are made up of base words together with all of their derived and inflected forms that can be understood by a learner without having to learn each form separately (Bauer & Nation, 1993:253) (see Section-2.6 p.16), and they were chosen as the source of words in Nation's Vocabulary Size Test (2009) because there is "increasing evidence that the word family is a psychologically real unit" (Nation & Beglar, 2007:10) that learners build their vocabulary upon. The word frequencies in Nation's Vocabulary Size Test (2009) were arranged according to the spoken section of the British National Corpus because it was felt that the written section of that corpus had content that was too formal (Nation & Beglar, 2007:10), and that the word frequencies arising from the spoken section would more closely represent the order in which the intended test-takers might learn the words (*ibid.*, pp.10-11). Each word in Nation's Vocabulary Size Test (2009) was accompanied by an example sentence that put the word in a non-defining context, which could help test takers know what part of speech the word belonged to, but would not help in understanding further meaning (*ibid.*, p.11). The different possible word definitions that were used for the multiple choice meaning options were carefully formulated to ensure they used easier words than the tested word itself (Lott, 1960:66), being achieved by using words from a more frequent word level (Nation & Beglar, 2007:11), and once again, the options had the tested words in non-defining contexts (*ibid.*, p.12). Finally, initial trials for Nation's Vocabulary Size Test (2009) showed that successful non-native English speaking (NNS) undergraduates studying at a university in an English speaking

country had a vocabulary size of 5000-6000 word families, while postgraduates of a similar background had a size of 9000 (*ibid.*, p.12).

With this background, Nation's Vocabulary Size Test (2009) was used to provide the target words for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning). Its arrangement of words in decreasing word frequency levels was a key feature, because it would serve to gauge 1 of the 3 aspects of partial vocabulary acquisition under investigation in this study, namely gauging partial vocabulary acquisition by testing vocabulary from different word frequency levels (see Section-3.4.2 above).

However, fewer words were used in this study's Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) when compared to those used in Nation's Vocabulary Size Test (2009), according to the following meticulously thought-out reasons and criteria. Firstly, the target words were only taken from up to the 10000-word level (i.e. not from 11000 to 14000-word levels) because it was thought that the participants in the study would be very unlikely to know words beyond this, just as Al-Homoud & Schmitt (2009:390) assumed in their study of Saudi university students in Riyadh. Also, instead of 10 words per level, only 5 were chosen making a total of 50 tested words (see Appendix-1 & Appendix-2 for the original and chosen words). It has been recommended that research, especially when carried out by teachers in real teaching contexts, should be practical and easy to implement, to make such research sustainable (Kluge, 1997:4). This was especially applicable to this study, because its real classroom setting (see Section-2.8.2 p.38) required that as little time as possible should be taken away from the teaching time of the normal course, and so short tests would be quicker and easier to administer. Another reason for avoiding long tests was the 'law of diminishing returns' (Schmitt, 1994:12) that sets in when participant fatigue leads to careless responses. Again, the use of shorter tests would minimize this problem, leading to a potentially better quality of responses.

A number of criteria were considered when choosing the target words (see Appendix-1 & Appendix-2 for the original and chosen words). Firstly, shorter words are generally considered to be easier to read than longer words (Nation & Coady, 1988:97), so

wherever it was reasonably possible, the words were chosen to have broadly similar lengths in terms of number of letters, leading to for example, the very short word ‘jug’ in the 3000-word level being omitted. Secondly, it is thought that words with more abstract meanings are harder to understand than those with more concrete meanings (Waring & Nation, 2004:101; Nation & Coady, 1988:97), and so for example ‘gimmick’ in the 7000-word level was not used because it was deemed to have a meaning that was too abstract. In addition, words that are context-specific or culture-specific may be very difficult for many learners to understand (Pellicer-Sanchez & Schmitt, 2010:37) if they are unfamiliar with that context or culture. Hence ‘bacterium’ in the 5000-word level was not used because it was thought to be too specific to a context (biology) that the students (who were all studying technical majors) would not be familiar with, while ‘vocabulary’ and ‘input’ in the 4000-word level were omitted because they were thought to be too easy, given their repeated use in very familiar contexts, such as in their English classes and technical classes. Similarly, words like ‘pub’ in the 2000-word level and ‘ruck’ (which is predominantly used for the sport of rugby) in the 10000-word level were not used, because they were thought to be too specific to the English culture that the students of this study would not have been exposed to.

The final criteria considered when choosing the target words (see Appendix-1 & Appendix-2 for the original and chosen words), was the class of the word, with verbs being thought to be harder to understand than nouns because of their different forms (Waring & Takaki, 2003:136), and because verbs occur much less frequently than nouns, as was found by Brezina & Gablasova (2013:5-8) when investigating the contents of 4 corpora that contained many billions of words. Hence, of the 50 chosen words, 26 were nouns, 14 were verbs and 10 were adjectives/complements, which was an arguably better representation of word classes than in some surveyed studies, such as Pellicer-Sanchez & Schmitt (2010:38) who almost entirely used nouns, Pigada & Schmitt (2006:9) who used roughly equal numbers of only nouns and verbs, and Waring & Takaki (2003:136) who used only nouns and adjectives.

It should be acknowledged that it is quite possible that the chosen target words may not be as well aligned to the original test constructs as the full cohort of original words, and hence they would not be as good for determining vocabulary size, as envisaged by those who made the original Vocabulary Size Test (Nation, 2009). However in this study, the tests with the chosen target words were not used to estimate vocabulary size. Instead, they were simply used to gauge vocabulary acquisition during the ER programme, by using them as pretests and posttests that contained words that were already conveniently arranged according to the different word frequency levels.

3.4.4.1 *Details of Test-1 (Spelling)*

Test-1 (see Appendix-3 & Appendix-6) was designed to assess receptive knowledge of spelling through multiple choice questions (MCQs), and was combined with Test-2, which was designed to assess productive knowledge of meaning through open translation (see Section-3.4.4.2 p.84 and Appendix-3 & Appendix-6). These two were combined onto one paper to simply reduce test administration time, with this combination being feasible because neither test could give clues to each other's answers.

Spelling in English is problematic to many learners because of its 'illogicality... [and] ...arbitrary nature' (Moody, 1974:318) when compared to what could be plausible based on pronunciation alone. Hence, some of the options for the answers of the MCQs were based on the phonology of the word, while others were based on the spelling of words that sounded similar or were spelled similarly, and others were based on wrong spellings that the researcher had encountered previously from students in this context. In addition, it has been advised that wrong options should not unduly stand out (Schmitt (1994:12), and so plausible spellings that followed English spelling conventions were mainly used for the options, as recommended and done by Waring & Takaki (2003:136).

As described in Section-3.1.1 p.42, Stage-1 of the investigation was the Initial Pilot Study in which all the data collection instruments were trialled with an Expert focus

group of 7 English teaching colleagues and a Student focus group of 4 students. All of these were involved in piloting Test-1 (Spelling), and they gave a great deal of qualitative feedback through the Supplementary Questioning (Section-3.8 p.117) that resulted in changes being made to some of the items (see Appendix-5). For example, Item-35 Option-A had the spelling 'yogurt' which was changed to 'yogort' after it was pointed out that the former is also another correct spelling for the target word of 'yoghurt'. In addition, there was widespread agreement that the 6-option Likert scale used to gauge level of certainty was too detailed. The '0' option (representing a response that was a blind guess) and the '5' option (representing a response the participant was absolutely sure about was correct) were clear, but many were unsure about how the 4 intermediate numbers exactly differed from each other. Hence this was changed in the revised version of all the 4 tests to a 4 point scale, from '0' to '3', with '3' representing a response the participant was absolutely sure was correct. Figure-1 below shows the first part of the revised version that is given fully in Appendix-6.

Figure-1: The First Few Questions of the Revised Versions of Test-1 (Spelling) & Test-2 (Open Translation) Combined [See Appendix-6 for the full version]

Circle the letter [A, B, C or D] of the sentence which spells the bold underlined word in the correct way.

Write CLEARLY a word or sentence in ARABIC that explains the meaning of the English word

هات كلمة عربية أو جملة عربية تفسر معنى الكلمة الإنجليزية الصحيحة واكتبها بخط نسخ واضح

Circle the number that indicates how sure you were of your answer

ضع دائرة حول تقييمك لدرجة التأكد من جوابك

E.g. Circle '0' if you have no idea and are just blindly guessing

مثلا: ضع دائرة حول "0" لو لم يكن لديك أية فكرة

E.g. Circle '3' if you feel absolutely sure

مثلا: ضع دائرة حول "3" لو كنت متأكدًا جدًا

1	A They <u>siw</u> it.	B They <u>saw</u> it.	C They <u>sur</u> it.	D They <u>sor</u> it.	3	2	1	0
					3	2	1	0
2	A They have a lot of <u>tim</u> .	B They have a lot of <u>tyme</u> .	C They have a lot of <u>time</u> .	D They have a lot of <u>tym</u> .	3	2	1	0
					3	2	1	0
3	A We are <u>poor</u> .	B We are <u>poar</u> .	C We are <u>por</u> .	D We are <u>pur</u> .	3	2	1	0
					3	2	1	0

This revised version (Appendix-6 and Figure-1 above) was fully trialled during the Main Pilot Study with all of its 52 participants. The results were analysed statistically and satisfied the relevant checks, as detailed below in Section-3.4.7 p.90 and Appendix-

15. This version-2 of Test-1 (Spelling) (see Appendix-6) was then used in the Main Study.

3.4.4.2 *Details of Test-2 (Translation)*

Test-2 (based on Nation, 2009) (see Appendix-3 & Appendix-6) required the participants to give an Arabic translation for each English word in order to gauge their productive knowledge of meaning. It was combined with Test-1 (Spelling) on the same paper to reduce test administration time, with this combination being feasible because neither test could give clues to each other's answers (see Section-3.4.4.1 p.82).

During the Initial Pilot Study, the 2 native Arabic-speaking colleagues of the Expert focus group, and the whole Student focus group were involved. It was envisaged that a productive test (giving the meanings of English words through open Arabic translations) would be more difficult than a receptive test for meaning, as found by Ishii & Schmitt (2009:12), and so the feedback from the participants was particularly important.

A basic feature of a second language test should be that proficient participants should be able to score almost perfect scores (Schmitt et al., 2001:65), and this was the case with the 2 native Arabic-speaking colleagues of the Expert focus group. Their responses also provided an initial bank of correct translation answers, against which the responses of the Student focus group was checked. The 2 native Arabic-speaking colleagues were further consulted about seemingly incorrect responses from the Student focus group, yielding valuable information, such as Arabic words that were correct according to the local dialect the students spoke in, but were unfamiliar to the researcher because of his knowledge of predominantly academic Arabic. Hence, the input of the native Arabic-speaking colleagues was important in giving another angle to the accuracy of the answers, noting that researchers such as Hughes & Lascaratou (1982:175-177) and James (1977:118) have found significant differences in how NS and NNS teachers judge the accuracy of students' answers. Further reference was made to native Arabic-speaking colleagues during the Main Pilot Study, and so the bank of correct translation

answers was expanded, with the final version for the Main Study being given in Appendix-7.

Overall, the feedback received during the Initial Pilot Study required no major changes to be made to the test, except the rationalizing of Likert scale options used to gauge the level of certainty of each response, as explained above in Section-3.4.4.1 p.82 for Test-1 (Spelling). Figure-1 in Section-3.4.4.1 above shows the first part of the revised version that is given fully in Appendix-6.

This revised version (Appendix-6 and Figure-1 in Section-3.4.4.1) was fully trialled during the Main Pilot Study with all of its 52 participants. During this, the bank of correct translation answers (see Appendix-7) was expanded to include new correct responses provided by the larger pool of participants (n=52), with consultations about these continuing with native Arabic speaking colleagues. The results were analysed statistically and satisfied the relevant checks, as detailed below in Section-3.4.7 p.90 and Appendix-15. This version-2 of Test-2 (Translation) (see Appendix-6) was then used in the Main Study.

3.4.4.3 *Details of Test-3 (Meaning)*

Test-3 (Meaning) (see Appendix-4) was designed to assess receptive knowledge of meaning through multiple choice questions (MCQs), used the same 50 words as Test-1 (Spelling) and Test-2 (Open Translation), and had the same MCQ options as in Nation's Vocabulary Size Test (2009). The test was piloted during the Initial Pilot Study, with 6 of the 7 colleagues of the Expert focus group, and the whole Student focus group being involved. This yielded very rich feedback, with perhaps more comments and suggestions than for any other test, and hence resulted in 6 of the test items being modified (see Appendix-5). For example, Item-15 'lonesome: He felt lonesome' Option-C had the meaning 'lonely' which was changed to 'without a friend' simply because the former ('lonely') was deemed to be too similar to the item word ('lonesome'). Care was also taken to ensure that the revised MCQ options had vocabulary that was from a more frequent word level than that of the item word itself

(see Appendix-5 for details). In addition, the Likert scale options (used to gauge the level of certainty of each response) were rationalised, as explained in Section-3.4.4.1 p.82 for Test-1. Figure-2 below shows the first part of the revised version that is given fully in Appendix-8.

Figure-2: The First Few Questions of the Revised Version of Test-3 (Meaning) [See Appendix-8 for the full version]

Circle the letter [A, B, C or D] which explains the meaning of the word in the most correct way.
 Circle the number that indicates how sure you were of your answer
 E.g. Circle '0' if you have no idea and are just blindly guessing
 E.g. Circle '3' if you feel absolutely sure

ضع دائرة حول تقييمك لدرجة التأكد من جوابك
 مثلاً: ضع دائرة حول "0" لو لم يكن لديك أية فكرة
 مثلاً: ضع دائرة حول "3" لو كنت متأكدًا جدًا

1 see: They saw it.	a. cut	b. waited for	c. looked at	d. started	3 2 1 0
2 time: They have a lot of time.	a. money	b. food	c. hours	d. friends	3 2 1 0
3 poor: We are poor.	a. have no money	b. feel happy	c. are very interested	d. do not like to work hard	3 2 1 0
4 jump: She tried to jump.	a. lie on top of the water	b. get off the ground suddenly	c. stop the car at the edge of the road	d. move very fast	3 2 1 0
5 basis: I don't understand the basis.	a. last part	b. words	c. road signs	d. main part	3 2 1 0

This revised version (Appendix-8 and Figure-2 above) was fully trialled during the Main Pilot Study with all of its 52 participants. The results were analysed statistically and satisfied the relevant checks, as detailed below in Section-3.4.7 p.90 and Appendix-15. This version-2 of Test-3 (Meaning) (see Appendix-8) was then used in the Main Study.

3.4.5 Test-4 (Meaning)

The final test, Test-4 (Meaning) (Appendix-9 and Appendix-10), was designed to assess receptive knowledge of meaning through multiple choice questions (MCQs), but was based on the Vocabulary Levels Test of Schmitt et al. (2001), which has been used in numerous studies (Al-Homoud & Schmitt, 2009:389). It tests the meaning of 30 words in each of 4 word frequency levels (the 2000, 3000, 5000 & 10000-word levels) and one additional group of 30 academic words, giving a total of 150 words. The layout of the questions was different to the usual MCQ layout of one item word with 4 different

meaning options to choose from. Instead, 3 defined meanings were to be matched with 3 words from a possible 6 (see Appendix-9), making the words the options instead of the definitions (Schmitt et al., 2001:59). Only nouns, verbs and adjectives were used, to reflect their dominance in English, and a ratio of 3:2:1 was used so that each section contained three noun clusters, two verb clusters and one adjective cluster (*ibid.*, p.58). Each option word for each cluster was also chosen to be from the most frequent word of its word family (see Section-2.6 p.16), to have very different meanings from each other, to have different orthographic forms from each other, and to be as different as possible to the words in the cluster definitions (*ibid.*, p.59).

The main difference between the original Vocabulary Levels Test (Schmitt et al., 2001) and Test-4 (meaning) (Appendix-9 and Appendix-10), was that Test-4 had instructions and examples so that the participants also gave a number to denote how certain they were of each answer, in order to gauge this aspect of partial vocabulary acquisition (see Section-3.4.2 above). The main differences between Test-3 and Test-4 were the layout, the quantity of words and the choice of words, noting that the Vocabulary Levels Test (Schmitt et al., 2001) (which is the basis of Test-4) is a diagnostic test that just looks at the 5 above-mentioned slices of a learner's vocabulary, while the Vocabulary Size Test (Nation & Beglar, 2007:10) (which is the basis of Test-3) is a proficiency test used to determine how much vocabulary learners know, looking at all the frequency levels from 1000 to 14000 (see Section-3.4.4). In addition, the Vocabulary Size Test (*ibid.*, p.11) has distractors among the options for the MCQs, which share elements of meaning with the correct answer option, and so could be harder than the Vocabulary Levels Test (Schmitt et al., 2001).

The test was piloted during the Initial Pilot Study with 6 of the 7 colleagues of the Expert focus group and the whole Student focus group. Particular attention was given to the results of the 5000-word and 10000-word frequency levels, because Al-Homoud & Schmitt (2009:390) chose not to use these sections when testing their participants who were also Saudi students of low proficiency. The results and feedback showed that, although the scores for the lower frequency levels were much less than the high frequency levels and the academic word level, the scores were enough to make them

worthy of inclusion. However, there was almost unanimous agreement from all the participants of the Expert and Student focus groups that the layout of the test was strange and confusing, because none of them had ever seen such a layout before for a multiple choice test. Hence Test-4 (Meaning) was redesigned to have a more familiar multiple choice format (version-2 – see Appendix-10), and Likert scales with only 4 options were used to gauge the certainty of the chosen answers (see Section-3.4.4.1 above for the reasons of this rationalization of options). Figure-3 below shows the first part of the revised version that is given fully in Appendix-10.

Figure-3: The First Few Questions of the Revised Versions of Test-4 (Meaning) [See Appendix-10 for the full version]

<p>Circle the letter [A, B, C, D, E or F] of the word which explains the sentence on the left in most correct way. ضع دائرة حول الكلمة الصحيحة Circle the number that indicates how sure you were of your answer ضع دائرة حول تقييمك لدرجة التأكد من جوابك E.g. Circle '0' if you have no idea and are just blindly guessing مثلاً: ضع دائرة حول "0" لو لم يكن لديك أية فكرة E.g. Circle '3' if you feel absolutely sure مثلاً: ضع دائرة حول "3" لو كنت متأكدًا جدًا</p>							
<p>2000 Word Level درجة التأكد</p>							
1 end or highest point	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
2 this moves a car	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
3 thing made to be like another	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
4 loud deep sound	A accident	B debt	C fortune	D pride	E roar	F thread	3 2 1 0
5 something you must pay	A accident	B debt	C fortune	D pride	E roar	F thread	3 2 1 0

A particular concern with this revised version of Test-4 was that to fit all the 150 items, the test became quite dense with small-sized font. Few complained about this aspect of the test, but in order to eliminate any chance of this problem, a plain sheet of horizontally aligned A4 paper was given to each student to ‘underline’ each question. This made the whole line that contained the target word and its answer options stand out from the rest, making it much clearer and easier to complete.

This revised version (Appendix-10 and Figure-3 above) was fully trialled during the Main Pilot Study with all of its 52 participants. The results were analysed statistically and satisfied the relevant checks, as detailed below in Section-3.4.7 p.90 and Appendix-15. This version-2 of Test-4 (Meaning) (see Appendix-10) was then used in the Main Study.

3.4.6 Administration of Vocabulary Test-1, Test-2, Test-3 & Test-4

As outlined in Section-3.1.2 Table-5 p.48, the pretests were administered during the start of the course, and the posttests near the end, but in a way that minimized disruption to normal teaching and reduced participant fatigue. This first aspect was very important given this study's important characteristic of being set in the normal learning environment of the students (see Section-2.8.2 p.38), while the second aspect was important to ensure that the students could give their full concentration, because fatigue is often seen with long tests, which leads to poor responses (Schmitt, 1994:12). To achieve both of these, the tests had to be as quick as possible to complete. During the Initial Pilot Study and the Main Study (see Section-3.1.1 p.42) it was observed that Test-1 (Spelling) & Test-2 (Open Translation) combined typically needed 30 minutes to complete, Test-3 (Meaning) needed 20 minutes and Test-4 (Meaning) needed 1 hour (noting that Test-1, Test-2 & Test-3 were based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4 p.79, and Test-4 was based on the Vocabulary Levels Test (Schmitt et al., 2001). Hence Test-1, Test-2 and Test-3 were taken during the same lesson, with the combined Test-1 (Spelling) and Test-2 (Open Translation) being given first, because their answers could not help with Test-3 (Meaning). Test-4 (Meaning) was then taken during the lesson on the day after. A time limit of 1 hour and 40 minutes was given (i.e. the duration of the lesson), and in all cases during the Main Study, students were able to complete the day's test(s) within that time.

The monitor for all the tests was the researcher, and so the administration procedure was maintained during the pretests and the posttests. The students were all new to the college, and they were informed that the tests were for diagnostic purposes to see their current level of vocabulary knowledge. Hence they were told, as advised by Dörnyei (2003:41), not to think long over any item, but just give their natural responses, and that the results of the tests had no effect on their final course grades. The students were also reminded again of this 3 or 4 times while the tests went on. Some students were chosen randomly at the start of each test to read out each line of the written instructions and examples, both in English and Arabic, and the monitor (the researcher) explained them all again, but verbally and in Arabic. The students were instructed to raise their hands

once the first page of a test was completed, so that the monitor could check completion, and the monitor continued to check for the same throughout the test, also asking for clarification concerning interesting, unexpected or unclear answers. Papers were also checked once again at the end, while students were still present to give clarification.

The students were not informed of the future posttests, but there was a concern that some students may, even out of curiosity, actively research some words after the pretest, which could then affect the posttest results. However, it was readily observed that students did not do this, perhaps because they were nervous and somewhat disoriented during their first week in a new college in a new city. In addition, the many events of their first week at college, combined with the testing of different sets of vocabulary over two days (the 50 words of Test-1, Test-2 & Test-3 (all based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4 p.79) on one day, and the 150 different words of Test-4 (based on the Vocabulary Levels Test (Schmitt et al., 2001) as detailed in Section-3.4.5 p.86) on the next day) and followed by the immediate start of the course, served to reduce any particular attention to the words of the tests, with Pigada & Schmitt (2006:10-11) specifically recommending ‘distraction tasks’ to flush the target words from the participants’ immediate memory, and hence stop any explicit attention to them. During the Main Pilot Study, the students were also specifically asked at the start of each posttest whether they remembered any word in particular from the pretests, and in all cases the answers were negative. This formal procedure was also used during the Main Study with the experimental and control cohorts, and again no one reported any specific memory of the pretest items.

3.4.7 Statistical Analysis of the Main Pilot Test Results

As detailed above (Section-3.4.3, Table-11 p.77), in order to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36), 4 pretests and posttests of different aspects of vocabulary knowledge were developed during the Initial Pilot and the Main Pilot.

The tests were also designed to allow for partial vocabulary acquisition to be gauged in three ways (see above Section-3.4.2 p.75):

1. Testing different facets of word knowledge
2. Testing vocabulary from different levels of word frequency
3. Accounting for different levels of certainty for any given answer

These yielded a vast quantity of data, which needed a careful and consistent methodology of analysis to ensure the internal reliability of the research, whereby other independent researchers could re-analysis the same data to yield similar results (Nunan, 1992:14-17).

Hence first, the results were analysed separately for each of the tests, resulting in a separate analysis for each aspect of word knowledge:

1. Test-1 (Spelling) based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4 p.79,
2. Test-2 (Open Translation) based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4,
3. Test-3 (Meaning) based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4,
4. Test-4 (Meaning) based on the Vocabulary Levels Test (Schmitt et al., 2001) as detailed in Section-3.4.5 p.86.

Second, within each test, the results were analysed separately according to the different word frequency levels:

1. The results for each of Test-1 (Spelling), Test-2 (Open Translation) and Test-3 (Meaning) (all based on the Vocabulary Size Test, Nation, 2009, as detailed in Section-3.4.4 p.79) were split into the 1000-5000-word levels and the 6000-10000-word levels, with each frequency band having 25 words.
2. The results for Test-4 (Meaning) (based on the Vocabulary Levels Test, Schmitt et al., 2001, as detailed in Section-3.4.5 p.86) were split into the 2000-word, 3000-word, Academic-word, 5000-word, and 10000-word levels, with each of these 5 levels having 30 words.

Test-1, Test-2 & Test-3 were split into just 2 broad word frequency levels, because the participants in this study were generally lower-level (see the results of the Placement Tests above in Section-3.3 p.69) and the ER programme was relatively short (13.1 weeks - see Section-3.2.5 p.66), so it was anticipated that most vocabulary gains would be shown within the lower 1000-5000-word levels. This upper limit was chosen because the 5000-word level is needed for reading unsimplified material (Hirsh & Nation, 1992:689), and it was thought that few students would advance beyond this during the ER course. Test-4 (based on the Vocabulary Levels Test, Schmitt et al., 2001, as detailed in Section-3.4.5) was already split into its 5 word frequency levels. Hence for each of Test-1, Test-2 & Test-3, two separate sets of 25 words were analysed, and for Test-4, five separate sets of 30 words were analysed, with these numbers being comparable to those analysed in previous studies on vocabulary acquisition (see Section-3.1.2 p.45).

Third, the results were analysed separately for:

1. Correctness Alone
2. Correctness Factored by Level of Certainty
3. Certainty Alone.

Correctness Alone was analysed according to whether the answer was simply correct or not. This was obtained by assigning 1 to a correct answer, and 0 to an incorrect answer. This was analysed separately to enable comparisons to be made with the findings of other studies, most of which have only tested correctness, without considering level of certainty (see Section-3.4.2 p.75). There was one single exception to this: if a correct answer was accompanied by a level of certainty of 0 (i.e. the entry was a blind guess, but just happened to be correct), a score of 0 was assigned. This was chosen to be an effective way of preventing blind guesses (even if unintentionally given) from distorting the results, noting this problem was poorly addressed in the tests used in previous surveyed studies, as detailed in Section-3.4.2. Hence for the example entries in Figure-4 below, Item-1 would score '1', Item-2 '0' and Item-3 '0'.

Figure-4: Example Entries for the First Three Items of Test-3 (Meaning) [See Appendix-8]

1	see: They saw it.	a. cut	b. waited for	c. looked at	d. started	3	2	0
2	time: They have a lot of time.	a. money	b. food	c. hours	d. friends	3	2	1
3	poor: We are poor.	a. have no money	b. feel happy	c. are very interested	d. do not like to work hard	3	2	0

Correctness Factored by Level of Certainty was analysed in order to weight a correct answer according to how certain the student was about it. This was obtained by simply multiplying the correctness score with the level of certainty. For example, a student who obtained a correct answer, but recorded a level of certainty for that answer of 1 out of 3 would score $1 \times 1/3 = 0.333$, and so for the example entries in Figure-4 above, Item-1 would score '0.667', Item-2 '0.000' and Item-3 '0.000'.

Certainty Alone was essential to gauge if weaker students improved in confidence, even if their answers remained incorrect, because Correctness Factored by Level of Certainty would always score 0.000 for a wrong answer, even if the student's confidence (i.e. level of certainty) was improving. For example, a student who recorded a level of certainty for an answer of 1 out of 3 would score $1/3 = 0.333$, even if incorrect, and so for the example entries in Figure-4 above, Item-1 would score '0.667', Item-2 '0.333' and Item-3 '0.000'.

Hence in summary, the analysis was separated into:

1. 4 separate tests, gauging 3 different aspects of word knowledge;
2. Each of the above contained a pretest and posttest;
3. Test-1 (Spelling), Test-2 (Open Translation) and Test-3 (Meaning) (all based on Nation, 2009) tested 50 words that were separated into 2 groups of word frequency;
4. Test-4 (Meaning) (based on Schmitt et al., 2001) tested 150 words that were separated into 5 groups of word frequency;
5. All of these were then further separated into Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone.

Therefore, a total of 66 sets of data were separately analysed, for each of the 52 participants of the Main Pilot, for all of the 200 word items of vocabulary, as shown in Figure-5 p.95 below & Appendix-15.

A large number of Excel spreadsheets were used to manipulate this vast quantity of data, and IBM's SPSS software (v.17 and higher) was used to process it to yield results for both the descriptive and inferential statistics. The descriptive statistics included means, standard deviations, medians, and quartiles, while the inferential statistics included Cronbach's alpha coefficient for mean inter-item correlation to check the internal consistency (Pallant, 2010:97) of the tests, and also included hypothesis testing to compare the pretests and posttests for significant gain. As seen below, the descriptive statistics showed that the data was not normally distributed, and so the non-parametric Wilcoxon Signed-Rank Test for repeated measures was used to compare the pretests and posttests (*ibid.*, p.230).

The results of the vocabulary tests during the Main Pilot are given in Appendix-15, and are summarised below in Figure-5. Further details below in Section-3.4.7.1 p.96 to Section-3.4.7.7 p.100 show that these tests satisfied the relevant statistical checks (see Appendix-6, Appendix-8 & Appendix-10), and hence there was confidence in using them during the Main Study.

Figure-5: Summary of Vocabulary Test Results during the Main Pilot [See Appendix-15]

TEST-1 Based on NATION (2009)
SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.805	10.0	0.777	11.0	0.077	-1.77	0.17	1.0	10%
	Correctness x Certainty	0.852	8.2	0.846	9.2	0.008	-2.67	0.26	1.0	12%
	Certainty only	0.923	13.5	0.919	14.0	0.060	-1.88	0.18	0.5	4%
6000-10000wd 25 items	Correctness only	0.749	5.0	0.798	5.0	0.325	-0.99	0.10	0.0	0%
	Correctness x Certainty	0.733	2.8	0.869	3.0	0.341	-0.95	0.09	0.2	6%
	Certainty only	0.927	5.8	0.934	6.8	0.188	-1.32	0.13	1.0	17%

TEST-2 Based on NATION (2009)
TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.821	6.0	0.853	6.5	0.000	-3.69	0.36	0.5	8%
	Correctness x Certainty	0.846	5.0	0.872	6.3	0.000	-3.65	0.36	1.3	27%
	Certainty only	0.870	8.0	0.889	10.0	0.006	-2.76	0.27	2.0	25%
6000-10000wd 25 items	Correctness only	0.426	0.0	0.899	0.0	0.019	-2.35	0.23	0.0	Pretest = 0.0
	Correctness x Certainty	0.487	0.0	0.924	0.0	0.005	-2.83	0.28	0.0	Pretest = 0.0
	Certainty only	0.829	1.3	0.905	2.2	0.013	-2.48	0.24	0.8	63%

TEST-3 Based on NATION (2009)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.806	8.5	0.815	9.0	0.011	-2.55	0.25	0.5	6%
	Correctness x Certainty	0.858	6.3	0.862	7.3	0.000	-3.88	0.38	1.0	16%
	Certainty only	0.931	12.2	0.933	14.0	0.002	-3.08	0.30	1.8	15%
6000-10000wd 25 items	Correctness only	0.712	3.0	0.802	3.0	0.687	-0.40	0.04	0.0	0%
	Correctness x Certainty	0.744	1.7	0.893	1.8	0.207	-1.26	0.12	0.2	10%
	Certainty only	0.906	6.7	0.947	5.3	0.687	-0.40	0.04	-1.3	-20%

TEST-4 Based on SCHMITT ET AL (2001)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.913	7.5	0.918	8.0	0.009	-2.61	0.26	0.5	7%
	Correctness x Certainty	0.939	5.2	0.940	6.3	0.001	-3.34	0.33	1.2	23%
	Certainty only	0.961	14.2	0.956	15.0	0.067	-1.83	0.18	0.8	6%
3000wd 30 items	Correctness only	0.868	4.0	0.890	5.0	0.146	-1.45	0.14	1.0	25%
	Correctness x Certainty	0.892	3.0	0.924	3.3	0.029	-2.18	0.21	0.3	11%
	Certainty only	0.945	9.3	0.944	9.0	0.352	-0.93	0.09	-0.3	-4%
Academic wd 30 items	Correctness only	0.901	4.0	0.911	6.0	0.003	-3.01	0.30	2.0	50%
	Correctness x Certainty	0.923	2.8	0.933	4.3	0.001	-3.41	0.33	1.5	53%
	Certainty only	0.945	9.2	0.956	8.8	0.042	-2.03	0.20	-0.3	-4%
5000wd 30 items	Correctness only	0.799	2.0	0.813	3.0	0.052	-1.94	0.19	1.0	50%
	Correctness x Certainty	0.808	1.0	0.871	1.7	0.026	-2.23	0.22	0.7	67%
	Certainty only	0.936	5.0	0.947	4.7	0.231	-1.20	0.12	-0.3	-7%
10000wd 30 items	Correctness only	0.715	0.0	0.717	1.0	0.597	-0.53	0.05	1.0	Pretest = 0.0
	Correctness x Certainty	0.741	0.0	0.789	0.3	0.799	-0.26	0.03	0.3	Pretest = 0.0
	Certainty only	0.937	2.3	0.954	2.0	0.646	-0.46	0.05	-0.3	-14%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:
<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill. Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill; Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{(N_{\text{pretest}} + N_{\text{posttest}})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = ((posttest score – pretest score) / pretest score) x 100. Note if pretest score = 0.0, then Percentage Change cannot be calculated.

3.4.7.1 *Internal Consistency*

Cronbach's alpha coefficients were used to gauge inter-item correlation, and hence the internal consistency of the tests. The results for the Main Pilot (see Figure-5 p.95 & Appendix-15) demonstrated internal consistency between the items of the tests, with $\alpha > 0.7$ (Pallant, 2010:97) and these are highlighted in green. In many cases, the coefficients were considerably higher than 0.7, showing good internal consistency (*ibid.*, p.101). For example, $\alpha > 0.8$ for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) that were all based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4 p.79, and $\alpha > 0.9$ for Test-4 (Meaning) that was based on the Vocabulary Levels Test (Schmitt et al., 2001) as detailed in Section-3.4.5 p.86. Cronbach's alpha coefficients were also recalculated when each question was removed in turn, but this did not result in any substantial increase in the values.

Cronbach's alpha coefficient is sensitive to the number of items tested (Pallant, 2010:97), which could explain why the sections in Test-4 ($n=30$) had higher values than the sections in the other tests ($n=25$).

For just 2 of the 66 separately analysed sets of data, poor internal consistency of test items was found with $\alpha < 0.7$ (Pallant, 2010:97). These are highlighted in red in Figure-5 & Appendix-15, and they were for the pretest of Test-2 (Open translation) in the 6000-10000-word level frequency band. The reason could have been because open translation of these much less frequently occurring words was possibly very difficult for the participants in the beginning (i.e. during the pretest), but as they progressed during the duration of course, the students were able to attempt the questions with more confidence during the posttest, which did then yield better internal consistency.

3.4.7.2 *Results for the Different Aspects of Word Knowledge*

The results given above in Figure-5 & Appendix-15 between the different tests may tentatively suggest that with the tests that were based on the Vocabulary Size Test (Nation, 2009, as detailed in Section-3.4.4 p.79), the scores for Test-1 (Spelling) were

higher than Test-3 (Meaning), which in turn were higher than Test-2 (Open Translation). For example, the Median pretest scores for Correctness Alone for the 1000-5000-word levels were: 10.0/25 for Test-1 which was more than 8.5/25 for Test-3 which in turn was more than 6.0/25 for Test-2.

The Median pretest scores for Correctness Alone in Test-4 (Meaning) that was based on the Vocabulary Levels Test (Schmitt et al., 2001, as detailed in Section-3.4.5 p.86) were 7.5/30 for the 2000-word level, 4.0/30 for the 3000-word level, and 2.0/30 for the 5000-word level. This could suggest that Test-4 is approximately placed between Test-3 and Test-2 for those comparable word frequency levels.

Hence it could be cautiously implied that the results for Test-1 (Spelling) were higher than Test-3 (Meaning) which were more than Test-4 (Meaning) which were in turn more than Test-2 (Open Translation). This could indicate that the participants had more receptive knowledge of spelling and meaning than productive knowledge of meaning.

3.4.7.3 Results for the Different Word Frequency Levels

The results in Figure-5 above & Appendix-15 between the different word frequency levels suggest that for Test-1 (Spelling), Test-2 (Open Translation) and Test-3 (Meaning), the scores for the 1000-5000-word frequency levels were much higher than those for the 6000-10000-word levels. For example, the Median pretest scores for Test-1 for Correctness Alone were 10.0/25 for the 1000-5000-word levels which was double 5.0/25 for the 6000-10000-word levels, although in both cases the figures are still modest.

For Test-4 (Meaning) the scores in the 2000-word and Academic-word levels seemed to be higher than the other levels. For example, for Correctness Alone, the Median pretest scores for Test-4 were 7.5/30 for the 2000-word level > 4.0/30 for the Academic-word level > 2.0/30 for the 5000-word level. This was possibly expected, given that it was more likely that the participants would have better knowledge of words that appeared

more frequently in both general English and academic English contexts. Once again, it was observed that the scores were modest, even in the more frequent word levels.

Hence overall it may tentatively seem that for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning), the results for the 1000-5000-word levels were much higher than those for the 6000-10000-word levels. As for Test-4 (Meaning), the results may suggest that the 2000-word level had higher scores than the Academic-word level which were higher than the 3000-word level which were more than the 5000-word level which were in turn more than the 10000-word level. This could indicate that the participants had much better knowledge of high frequency general and academic words than of the lower frequency 5000-word & 10000-word levels, which are usually required for reading unsimplified texts (Hirsh & Nation, 1992:689).

3.4.7.4 *Results involving Levels of Certainty*

The results in Figure-5 p.95 above & Appendix-15 involving different levels of certainty showed that Correctness Alone was always higher than Correctness Factored by Level of Certainty (as should have been the case – see Section-3.4.7 p.90 above), while the scores seemed to be highest for Certainty Alone. For example, the Median pretest scores for Test-1 (Spelling) for the 1000-5000-word levels were 8.0/25 for Certainty Alone which was more than 6.0/25 for Correctness Alone which was higher than 5.0/25 for Correctness Factored by Level of Certainty.

The small difference between Correctness Alone and Correctness Factored by Level of Certainty may indicate that the students were fairly sure about most of their correct answers. The higher Certainty Alone scores could indicate that the students were fairly certain about a significant number of words, but were still getting them wrong.

3.4.7.5 *Distribution of Data*

The descriptive statistics are given in Appendix-15 and they show that, in general the standard deviations were large in comparison to the means, indicating a large spread of

results, and this was more pronounced in the less frequent word levels. For example, the pretest scores for Test-1 (Spelling) for Correctness Alone were Mean=6.4/25 & Standard Deviation=3.6 at the 1000-5000-word levels, compared to Mean=0.6/25 & Standard Deviation=1.0 at the 6000-10000-word levels.

The descriptive statistics also showed that the vast majority of the data was significantly skewed. For example, the pretest scores for Test-4 (Meaning) for Correctness Alone at the 2000-word level were Mean=9.2/30, Lower Quartile=4.0/30, Median=7.5/30 & Upper Quartile=12.0/30. This skew was more pronounced in the less frequent word levels. For example, the posttest scores for Test-1 for Correctness Alone at the 6000-10000-word levels were Mean=1.2/25, Lower Quartile=0.0/25, Median=0.0/25 & Upper Quartile=1.0/25.

Overall, the descriptive statistics showed that the vast bulk of the data was not normally distributed, and so the non-parametric Wilcoxon Signed-Rank Test for repeated measures was used to compare the pretests and posttests (Pallant, 2010:230) for statistically significant vocabulary gain.

3.4.7.6 *Vocabulary Gain between Pretests and Posttests*

As detailed above, the descriptive statistics showed that the data was not normally distributed, and so the hypothesis testing used to compare the pretests and posttests for statistically significant gain, was the non-parametric Wilcoxon Signed-Rank Test for repeated measures (Pallant, 2010:230).

These results in Figure-5 p.95 above & Appendix-15 showed that although there was an increase in the median score in almost every case of the 33 pretest-posttest pairs, this increase was statistically significant ($p < 0.05$) in the more frequent 1000-5000-word levels of Test-2 (Open Translation) & Test-3 (Meaning) that were based on the Vocabulary Size Test of Nation (2009) detailed in Section-3.4.4 p.79, as well as in the 2000-word and Academic-word levels of Test-4 (Meaning) that was based on the Vocabulary Levels Test of Schmitt et al. (2001) detailed in Section-3.4.5 p.86.

Furthermore, the significant gains in Test-4 (Meaning) were generally the largest, although in all the tests these increases were very modest in terms of absolute number of words gained.

For example, the Wilcoxon Signed-Rank Test for repeated measures revealed statistically significant vocabulary pretest-posttest gains during the ER programme for:

- Test-2 (Open Translation) Correctness Alone, 1000-5000-word levels: the score increased significantly, $z=-3.69$, $p=0.000$ to 3 decimal places, with a medium effect size ($r=0.36$), from pretest Median=6.0/25 to posttest Median=6.5/25;
- Test-3 (Meaning) Correctness Alone, 1000-5000-word levels: the score increased significantly, $z=-2.55$, $p=0.011$, with a small effect size ($r=0.25$), from pretest Median=8.5/25 to posttest Median=9.0/25;
- Test-4 (Meaning) Correctness Alone, 2000-word level: the score increased significantly, $z=-2.61$, $p=0.009$, with a small effect size ($r=0.26$), from pretest Median=7.5/30 to posttest Median=8.0/30;
- Test-4 (Meaning) Correctness Alone, Academic-word level: the score increased significantly, $z=-3.01$, $p=0.003$, with a medium effect size ($r=0.30$), from pretest Median= 4.0/30 to posttest Median=6.0/30.

3.4.7.7 *Summary*

Overall, the detailed results above for the vocabulary tests during the Main Pilot gave confidence that the testing instruments were suitable for use during the Main Study. The internal consistency checks were satisfied for almost all the 66 separately analysed sets of data, although few statistically significant vocabulary gains were found, and even those were modest in absolute terms.

Hence, in order to encourage more vocabulary gain during the Main Study, it was decided to increase the number of in-class sustained silent reading sessions (as described earlier in Section-3.2.3 p.61), and to increase the duration of the ER programme (see Section-3.2.5 p.66).

3.5 METHODOLOGY OF THE END-OF-SEMESTER EXAM

All participants of the Main Pilot (n=52) and the Main Study (n=85) completed the institution's English Language End-of-Semester Exam during the examination weeks after the end of the course (see Section-3.1.2 Table-5 p.48). The results were important to see if the time taken up by the pretests, posttests and ER during the lessons of the participants could have adversely affected their final examination results. If so, it would be completely unacceptable given that the final grade was a big concern to the students on the course. Hence the results of the end-of-semester exams for the Main Pilot cohort was checked with the results of their batch, and the findings are presented here. The results of the Main Study Experimental and Control cohorts were also compared with those of their batch, and those findings are detailed in Section-4.3 p.149.

For reasons of confidentiality that were assured when taking permission to carry out this study, a copy of the test cannot be included here. However it can be said that the test was 1 hour and 30 minutes long, and it consisted of a short listening section, a short reading comprehension section, and a final much bigger section of multiple choice and gap fill questions entirely based on the course book. For the purposes of this study, only the reading comprehension and main course book based section were considered, representing 90 per cent of the overall mark of the exam for the students of the Main Study.

Excel spreadsheets and IBM's SPSS (v.17 and higher) were used to process the data to yield results for both the descriptive and inferential statistics (see Appendix-19). The descriptive statistics (Pallant, 2010:59-63) showed that the data was not normally distributed, and so the non-parametric Mann-Whitney U Test for 2 independent samples was used (*ibid.*, p.227), where $p < 0.05$ indicated significant difference between the groups (*ibid.*, pp.240-242), i.e. between the Main Pilot cohort and the whole of their batch.

The above-mentioned Mann–Whitney U tests (see Appendix-19) showed that the End-of-Semester Exam results were significantly higher for the Main Pilot cohort (Median=87%, n=52) than for its total batch (Median=82%, n=730), $U=15458$, $z=-2.24$, $p=0.025$, $r=0.08$, while it should be noted as detailed in Section-3.3 p.69 that there was no significant difference beforehand at the start of the project between the Placement Test results of the Main Pilot cohort (Median=41%, n=48) and its total batch (Median=40%, n=593), $U=13948$, $z=-0.23$, $p=0.818$. This demonstrated that the participants of the Main Pilot were not disadvantaged by the ER programme during the final exams, and if anything they could perhaps have actually benefitted from it, at least in terms of their final exam results, even though the ER programme offered no explicit exam preparation. This gave confidence that the methodology would also be suitable for the eventual Main Study, and this was confirmed when analysing those results (see Section-4.3 p.149).

3.6 METHODOLOGY FOR THE READING DIARIES

In order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), diaries were utilised to gather information about the students' reading. Diaries are particularly useful for investigating reading habits outside the class because such habits are impossible to observe directly (Benson, 2001:201), and so they - like book reports (Bamford, 1984:219-221), logs (Hyland, 2002:188) and journals (Farrell, 2001:28) - are used as an important introspective tool to provide insights into language exposure that would be otherwise difficult to obtain. The importance of this cannot be overstated, because some surveyed reading studies took place in L2-rich environments, and so the associated proficiency gains could have resulted more from the L2 exposure outside the class than the reading inside the class (Hafiz & Tudor, 1989:9).

The diaries used in this study were also an integral part of the ER programme, being recommended because they offer a way to monitor and encourage the large amounts of reading that ER requires (Nation, 2001:3; Nation, 2005b:13-14). They were also formally used to develop the students' reflexive skills. Reflection is a process whereby grounds and assumptions are questioned (Hammond, 2006:272) and information is examined from different perspectives to generate new insights and understandings (Richards, 2003:114), with the overall aim being to improve the process under consideration. Opportunities for reflection have been recommended for teachers (Farrell, 2001:23) and students (Cotterall, 2000:110-111) alike, because reflection personalizes both teaching and learning, and leads to individual goals for improvement, all of which is thought to improve motivation (Dörnyei & Csizer, 1998:215). Hence by using the diaries to encourage reflection, the students directly benefited from the study in more than just linguistic terms, as recommended by Chappelle & Duff (2003:167) and Paran (2008:469). Further details of how this was implemented can be found in Section-3.6.4 p.108 below.

The reading diaries were developed during the Initial Pilot and Main Pilot studies (Section-3.1.1 p.42). This data collection instrument was to be used throughout the

course, and so it was particularly important that it was ‘carefully constructed and piloted to ensure reliability, avoid ambiguity, and to achieve a balance between having sufficient data and not overburdening respondents’ (Hyland, 2002:166). Hence, the first version used during the Initial Pilot study was more open and exploratory in design (as advised by Dörnyei, 2003:31), with its findings and feedback being used to limit and rationalise the version used during the Main Pilot Study, which in turn led to minor changes to the final version used during the Main Study.

Once the Reading Diaries of the Main Study were completed by the participants, the vast quantity of information contained therein needed a precise and consistent procedure of analysis. This methodology of analysis is detailed in full at the start of the Results section for the Reading Diaries (see Section-4.4 p.151).

3.6.1 Development of Items

The item pool for the Reading Diaries was developed based on a hierarchy of relevance that was determined through the experience of the researcher, and through the diaries, logs and book reports described in the reviewed literature (as also recommended by Dörnyei, 2003:32), and then the chosen items were defined as clearly as possible to ensure construct validity (Nunan, 1992:15-16). The first version of the Reading Diary Sheet used during the Initial Pilot is given in Appendix-11. Each sheet had 3 identical sections, to allow learners to record 3 different kinds of reading per sheet, with more sheets being available as needed. What follows is a rationale of the less straightforward items.

Item-3 was the Language of the reading material, as included before in some previous studies such as Janopoulos (1986:764). The inclusion of this item helped to promote respect for the students’ L1, which is an important aspect of preserving self-esteem and culture in English learners (McCabe, 2005:1), but it was also included to investigate L1 reading, which is thought to have an important role in many aspects of L2 learning (Nation, 2003:1). For example, understanding a text is based in part on the learner’s background knowledge (Nation & Coady, 1988:102-104; Schmitt et al., 2011:38) of the

context, especially when the text contains many exophoric references (Cutting, 2008:8), and so a learner who reads widely in L1 may have a wide background knowledge that could assist when reading in L2 (Schmitt et al., 2011:29-30). Also, learners' attitudes to L1 reading have been found to transfer to their L2 reading (Yamashita, 2004:1-2), with this being particularly relevant to the context of the participants in this study who read little in L1.

Item-4 was for the Type of reading material, which was left open for this pilot version of the Reading Diary Sheet. It took more time for the Student focus group of the Initial Pilot to complete, but the responses were not 'strait-jacketed', as described by Hyland (2002:167), and instead a richness of responses was generated, some of which were not anticipated (as predicted by Dörnyei, 2003:47).

Item-7 was to record the Amount of reading, which was measured in terms of the time spent reading. This was used in preference to Yamashita's (2004:5) choice of using the average number of pages read, deeming the later to be extremely difficult for the participants to gauge with any kind of accuracy or consistency.

Item-6 and Item-9 were to record the Time and Place of the reading respectively. These were included to investigate the popular times and places for reading in the context of these students.

Item-8 and Item-14 were to record the Reason of the reading and its Importance respectively. These were considered important items to investigate the students' motivation for reading different kinds of material. Item-8 was also left open to attract as many responses as possible, in order to develop a pool of options in revised versions of the diary.

Item-10 and Item-11 were to record the students' perception of the Ease of the reading, and to record Dictionary use respectively. They were put together for comparative purposes, with it being envisaged that more difficult texts would require more dictionary use, as reported by Asraf & Ahmad (2003:98).

Item-12 and Item-13 were to record the reading's Enjoyment and Interest respectively. The difference between the two was explained verbally to the participants with an example of a news article that appeared in a lesson at the beginning of the study. This article described the gruesome murder of a young child, and it was the kind of story that was not commonly found in the Saudi newspapers. The article, which was very much the topic of conversation among the students at that time, was an example of reading that they did not at all find enjoyable, but nonetheless still found it interesting in terms how shocking it was.

During the Initial Pilot Study, it was readily observed that completing the daily diary sheets was taking up too much time. It must always be reminded that the study took place in a real classroom setting, and so it was important that as little time as possible was taken away from the main course. Hence, it was decided during the Main Pilot Study to use weekly diary sheets, which were developed further, resulting in the final version of the Reading Diary Sheet given in Appendix-12. Each sheet had 2 identical sections, to allow learners to record 2 different kinds of reading per sheet, with more sheets being available as needed.

The most important difference between this version and the previous version (see Appendix-11 & Appendix-12) was that Item-3 (Type of Reading Material) was arranged in the named options that were yielded from the equivalent open response item in the previous version during the Initial Pilot Study. Hence to most popular types of reading during that initial stage were found to be college books, lecture hand-outs, student notes, newspapers, magazines, story books with enfacing translations, graded readers, simplified books with many pictures, simplified books with few pictures, the internet, mobile phones, films with translation subtitles, and full novels. A final open option was included for other types of reading not included in the Item, and this was usually comics. Using distinct named options enabled the diary sheets to be filled-in quicker, and hence take less time away from the main course.

3.6.2 Layout of Content

The layout of the diary sheet (see Appendix-12) was designed to be as clear as possible, using a tabular layout with as low a text density as possible, as advised by Dörnyei (2003:19-20) for questionnaires, and being equally applicable here. The English used was as simple as possible, as it has been noted that slight changes in wording can often lead to misunderstandings and varied responses (*ibid.*, pp.32-33), and these are even more likely when instructions are presented in a foreign language (Benson, 2001:196) to lower-level learners (Chappelle & Duff, 2003:167). Furthermore, Arabic translations were provided, which were checked for accuracy and ease of understanding with the Arabic speaking participants of the Initial Pilot Study (see Section-3.1.1 p.42).

The continuous data of Item-5 (Duration of Reading) was presented as grouped options, based on the reading durations found during the pilot stages. This simplification made item completion easier and quicker, bearing in mind that students would not be able to remember exactly how long they spent reading a particular material, and it also allowed for easier analysis of the results.

With other items Likert scales were used to also make administration quicker (Dörnyei, 2003:40) and analysis easier, because these items such as Item-7 (Ease of Reading) (see Appendix-12), were subjective in nature, and so Likert scales offered a ‘simple, versatile, and reliable’ (*ibid.*, p.36) way of quantification. Instead of the normal left to right arrangement, the numbers were arranged from right to left, following the reading direction used in the students’ L1 of Arabic. It has also been recommended to reverse the scales of similar items to monitor for superficial responding (*ibid.*, p.40) when the participant is being careless or not concentrating, and this was done for Item-7 (Ease of Reading) and Item-8 (Dictionary Use) (see Appendix-12). In this situation, it was expected that easy reading would be accompanied by little dictionary use, but these two items’ scales were arranged in opposite directions. Hence if a student was carelessly making entries, he may tick the left hand side of Item-7, recording the material was very easy, but may also tick the left hand side of Item-8, seemingly contradicting himself by recording that he needed to frequently refer to a dictionary. This would be easily noticed

by the teacher (i.e. the researcher) who checked the diary sheets as they were being filled in, allowing for immediate follow-up questions to be posed to ascertain what the actual entry should have been. Furthermore, it is often recommended to have an even number of options to prevent the unhelpful middle option being chosen by less thoughtful respondents (Al-Homoud & Schmitt, 2009:391; Dörnyei, 2003:36-38), but in these items, an odd number of options was chosen (1 to 5), simply because in this case it was decided that the middle option was a valid choice that should be accounted for.

3.6.3 Limitations

Like with all methods of self-reporting, the data collected through the reading diaries would have an element of inaccuracy, being wholly reliant on the accuracy of the participants' memories (Benson, 2001:203). This difficulty in remembering all the material read during the week would probably result in an underestimate of the actual reading done, but this was unavoidable given the need to use weekly (instead of daily) diary sheets, in order to ensure as little course time as possible was used. On the other hand, students may want to simply please the teacher (i.e. the researcher) by exaggerating their reading (Dörnyei, 2003:22), so this was mitigated as much as possible by continually reassuring them that what they did or did not record would not 'threaten' them (as termed by Dörnyei, 2003:58-59) in any way, because the reading and the diaries would never be tested or considered for their final grades.

3.6.4 Procedure

As outlined in Section-3.1.2 Table-5 p.48, the reading diaries were used to record the weekly reading of the participants of the Main Study, from the end of Week-1 up to Week-13 inclusive. During the pilot studies, the diaries were completed at the start of each week. However, it was readily seen that the students' memories were compromised as a result of the immediately preceding weekend, which they often spent in their hometowns with their large extended families, which in turn led to very little reading during that time. Hence, during the Main Study, the diary sheets were usually completed on the last day of the week, so the students were able to have clearer memories of what

they had read just prior during the week. This was always done under the supervision of the researcher to reduce the potential of 'hasty' and 'careless' entries, as described by Dörnyei (2003:16), and to also give an opportunity to immediately follow up with questions about interesting or unclear entries, as recommended by Hyland (2002:166).

As also outlined in Section-3.1.2 Table-5 p.48, the reading diaries were formally used to share ideas and promote reflection amongst the participants. This was done twice during the course: once after 4 full weeks of reading diaries were recorded, and once again in Week-10 after the mid-semester break. During the first session, each student was given the reading diaries of another student, and then asked to read them and note down one aspect of reading that he thought was good, and one aspect he thought needed improving. Afterwards, the students shared what they noted, and the teacher (i.e. the researcher) constantly reminded the students that the good aspects they found in their friend's reading habits should be emulated, while the aspects that required improvement should be taken as advice for the other student. The second session was held in the week after the mid-semester break, with an immediate aim of helping the students to get back into reading after the break, during which their reading was considerably less. The procedure was similar to that during the first session, but this time the students were asked to compare the first 2 weeks of reading diaries with the latest 2 weeks, and to note one improvement and one aspect of deterioration. Once again, the students shared their ideas, giving and receiving advice that encouraged their reading further. Throughout the discussion, the teacher (i.e. the researcher) posed questions to prompt reflection and to encourage the students to make targets (as recommended by Ferris, 1997:331) to improve their reading.

These two sessions of reflection and discussion enabled the students to directly benefit from the study in more than just linguistic terms, as recommended by Chappelle & Duff (2003:167) and Paran (2008:469). Formal reflection has been praised for providing opportunities to set goals and targets (Cotterall, 2000:116), to build up trust between learners and teachers (Farrell, 2001:36), and to encourage participants to slow down to assess their progress due to meaningful reflection that requires time and opportunity (*ibid.*, p.25). All of this was encouraged through the reflection and discussion sessions.

It also enabled the normally ‘individual solitary act of reading’ (Brown, 2009:243) to transform into a social activity and a kind of peer review, through which critiquing and providing feedback could gradually develop and strengthen the students’ reading (see Hu, 2005:321-322 in the context of writing, but which can be equally applicable here).

As mentioned above in Section-3.4 p.72, once the procedure for the Main Study Reading Diaries was completed, the large amount of data that was generated needed to undergo a carefully thought-out process of analysis. This methodology of analysis is detailed in full at the start of the Results section for the Reading Diaries (see Section-4.4 p.151).

3.7 METHODOLOGY FOR THE SURVEY OF EXPOSURE TO ENGLISH LEARNING

Surveys are important for gaining indirect evidence of unobservable out-of-class learning (Benson, 2001:201; Hyland, 2002:166), and so in order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36) a survey was administered to gauge the students' exposure to sources of English beyond the ER programme and the English Language course, in an attempt to address the limitations of some previous studies, such as Hafiz & Tudor (1989:9), that did not account for such exposure.

Furthermore, the survey was formally used to develop the students' reflexive skills, just like the Reading Diaries were (see Section-3.6.4 p.108 above). Reflection is a practice in which bases and conventions are probed (Hammond, 2006:272), and information is scrutinised from different angles to give new perceptions and understandings (Richards, 2003:114), with the overall goal being to enhance the process at hand. Reflection, as described before in Section-3.6.4 p.108, has been recommended for teachers (Farrell, 2001:23) and students (Cotterall, 2000:110-111) because it personalizes both teaching and learning, and it leads to personalised goals for improvement, all of which is thought to improve motivation (Dörnyei & Csizer, 1998:215). Hence by using the survey to encourage reflection, the students directly benefit from the study in more than just linguistic terms, as recommended by Chappelle & Duff (2003:167) and Paran (2008:469). Further details of how this was implemented can be found in Section-3.7.4 p.115 below.

Once the Survey was completed for the Main Study, the large quantity of information contained therein needed a precise and consistent procedure of analysis. This is detailed in full at the start of the Results section for the Survey of Exposure to English Learning (see Section-4.5 p.156).

3.7.1 Development of Items

The survey was developed during the Initial Pilot and Main Pilot studies (Section-3.1.1 p.42). Like the reading diaries, it was ‘carefully constructed and piloted to ensure reliability, avoid ambiguity, and to achieve a balance between having sufficient data and not overburdening respondents’ (Hyland, 2002:166). During these stages, the content and format of the survey was amended, with the biggest single change being the addition of a section to record English used in the work place, which arose during the pilot stages when one student said that he had used a lot of English during a summer work placement. The final version of the Survey of Exposure to English Learning used during the Main Study is given in Appendix-13.

The items (see Appendix-13) were developed based on a hierarchy of relevance that was determined through the experience of the researcher, and through what was found in surveys described in the reviewed literature (as also recommended by Dörnyei, 2003:32), and then the chosen items were defined as clearly as possible to ensure construct validity (Nunan, 1992:15-16). What follows is a rationale of the items.

Item-1 was for the Age of the participant. Age is an important psychological factor in language learning (Benson, 2004:10) and has been included in surveys in some previous studies, such as Janopoulos (1986:764). The researcher has also noted that almost all the mature students at the college have tended to display better language learning skills. This could be because age is related to background knowledge, which is thought to aid text comprehension (Nation & Coady, 1988:102-104; Schmitt et al., 2011:38), especially when the text contains many exophoric references (Cutting, 2008:8). Hence, an older learner may have more background knowledge that can also assist his L2 reading (Schmitt et al., 2011:29-30).

Item-2 was for English studied Outside the Current College, be it prior or concurrent. The type of institute was asked, because the researcher has seen a marked difference in language learning between students from government schools and private schools. The country of study was asked, again because the researcher has seen in the past a marked

difference in language learning between students who have studied abroad compared to those who have only studied in Saudi Arabia. The amount of teaching in English was asked in order to consider the typical situation in which the teacher teaches English using L1 via a predominantly grammar translation method, while the amount of English used outside the class was asked to investigate the differences between those who sufficed with using English during their lessons, and those who used English beyond their classroom environments.

Item-3 was for External English Exam results. This was included because the researcher had found before that the few students who had taken international exams like IELTS, or formalized English placement tests for jobs at large companies, have displayed stronger language learning abilities than others.

Item-4 to Item-8 related to English exposure during non-study related activities, such as travel, television, videos, video clips and radio, all of which can greatly enhance their language learning. In Item-4, the purpose of the journey was asked because travelling for pleasure and travelling for formal study may have different effects. The other items relating to watching and listening to English media could also have an effect on the books they choose to read (Hafiz & Tudor, 1989:10).

Item-9 related to English used at Work. A few of the students had worked at large companies that used English as the formal language of communication, and the researcher observed that they were noticeably stronger in their English than other students.

3.7.2 Layout of Content

The layout of the survey (see Appendix-13) was designed to be as clear as possible, using a tabular layout with as low text density as possible, as advised by Dörnyei (2003:19-20) for questionnaires. The English used was as simple as possible, as it has been noted that slight changes to wording can often lead to misunderstandings and varied responses (*ibid.*, pp.32-33), and these are even more likely when instructions are

presented in a foreign language (Benson, 2001:196) to lower-level learners (Chappelle & Duff, 2003:167). Furthermore, Arabic translations were provided, which were checked for accuracy and ease of understanding with the Arabic speaking participants of the Initial Pilot Study (see Section-3.1.1 p.42). The survey was short, being just 2 sides and taking the participants about 20 minutes to complete, following the advice of Dörnyei (2003:18) who recommended less than 4 sides and less than 30 minutes, and most items were close ended to also make them quicker to complete and easier to analyse, as recommended by Dörnyei (*ibid.*, p.35).

The continuous data of Item-5 to Item-8 (relating to the amount and frequency of watching and listening to English) (see Appendix-13) was presented as grouped options, based on what was found during the pilot stages. This simplification made item completion easier and quicker, bearing in mind that students would not be able to remember the exact figures for these, and it also allowed for easier analysis of the results. It has also been recommended to reverse the scales of similar items to monitor for superficial responding (Dörnyei, 2003:40) when the participant is being careless or not concentrating, and this was done for these 4 items (see Appendix-13). Hence for Item-5 and Item-6, it was expected the frequency and amount of watching of English films would roughly go hand in hand, so these two items' scales were arranged in opposite directions. Hence if a student was carelessly making entries, he may tick the left hand side of Item-5, recording a high frequency, but may also tick the left hand side of Item-6, seemingly contradicting himself by recording a small amount. This would be easily noticed by the teacher (i.e. the researcher) who checked the surveys as they were being filled in, allowing for immediate follow-up questions to be posed to ascertain what the actual entry should have been.

3.7.3 Limitations

Once again, like with all methods of self-reporting, the data collected through the surveys would have an element of inaccuracy, being wholly reliant on the accuracy of the participants' memories (Benson, 2001:203). This difficulty in remembering all their

prior exposure to English would probably result in an underestimate of the actual amount, but this was unavoidable with this retrospective method of data collection.

3.7.4 Procedure

As outlined in Section-3.1.2 Table-5 p.48, the Survey of Exposure to English Learning (Appendix-13) was administered during Week-11 of the course, and was done under the supervision of the researcher, to reduce the possibility of hasty or careless entries (as described by Dörnyei, 2003:16), and to also give an opportunity to immediately follow up with questions about interesting or unclear entries, as recommended by Hyland (2002:166).

As also outlined in Section-3.1.2 Table-5, the Survey of Exposure to English Learning (Appendix-13) was formally used to share ideas and promote reflection amongst the participants. This was done once in Week-12 after the survey was administered. During this session, each student was given the survey of another student, and then asked to read them and note down one aspect that he thought was good, and one aspect he thought needed improving. Afterwards, the students shared what they noted, and the teacher (i.e. the researcher) constantly reminded the students that the good aspects they found in their friend's reading habits should be emulated, while the aspects that required improvement should be taken as advice for the other student. Throughout the discussion, the teacher (i.e. the researcher) posed questions to prompt reflection and to encourage the students to make targets (as recommended by Ferris, 1997:331) to improve their reading.

This session of reflection and discussion enabled the students to directly benefit from the study in more than just linguistic terms, as recommended by Chappelle & Duff (2003:167) and Paran (2008:469). Formal reflection has been praised for providing opportunities to set goals and targets (Cotterall, 2000:116), to build up trust between learners and teachers (Farrell, 2001:36), and to encourage participants to slow down to assess their progress due to meaningful reflection that requires time and opportunity (Farrell, 2001:25). All of this was encouraged through the reflection and discussion

session. It also became a kind of peer review, through which critiquing and providing feedback could gradually develop and strengthen the students' ability to expose themselves to English beyond the classroom setting (see Hu, 2005:321-322 in the context of writing, but which can be equally applicable here).

As mentioned above in Section-3.7 p.111, once the procedure for the Survey during the Main Study was completed, the large amount of data that resulted needed to undergo a carefully thought-out process of analysis. This methodology of analysis is detailed in full at the start of the Results section for the Survey of Exposure to English Learning (see Section-4.5 p.156).

3.8 METHODOLOGY FOR THE SUPPLEMENTARY QUESTIONING

In order to address both RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), supplementary questions were posed throughout the study to the participants, and their responses were discussed to gain a deeper understanding of what they recorded in the data collection instruments, as recommended by Hyland (2002:166) and many others. The questions were posed in a systematic manner when checking entries in the data collection instruments, and at other times they were posed to immediately clarify strange or interesting entries. These clarifications were immediately written down on that particular student's data collection instrument, and also in the Research Journal (see Section-3.10 p.129 below).

The immediate nature of the questions ensured that the answers were as accurate as possible (as pointed out by Benson, 2001:203), and the informal nature allowed the students to be as relaxed as possible when answering. All of this also made the questioning during the administration of the ER programme and the data collection instruments more practical and less time consuming, which was an important consideration during this study that was set in a real teaching context.

In addition, once the results of the Main Study were available, further In-Depth Interviews were administered with a selection of students, in order to delve deeper into the potential factors that may have moulded the nature of the results (see Section-3.9 p.118 below). Hence, in this study a combination of immediate supplementary questioning and subsequent in-depth interviewing was used, combining the methods used in previously surveyed studies on ER and vocabulary, such as Al-Homoud & Schmitt (2009:395) who used only informal questioning with all their many participants, and Pellicer-Sanchez & Schmitt (2010:40) and Pigada & Schmitt (2006:10) who used only in-depth interviews with their few subjects.

3.9 METHODOLOGY FOR THE IN-DEPTH INTERVIEWS

In order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36) in-depth interviews were also conducted. Interviews are in many cultures ‘a natural and socially acceptable way of collecting information’ (Dörnyei, 2007:143), and so they are widely used as research tools in applied linguistics (Nunan, 1992:149). They are commonly used in qualitative research traditions (Dörnyei, 2007:134) such as ethnography, grounded theory, case studies and action research (Richards, 2003:13), but they have also been used in quantitative research (Spears, 1995:181), and in studies that have tested vocabulary acquisition. For example, Schmitt et al. (2001:57) used interviews to validate the summative scores in the vocabulary levels test they were developing, and Pellicer-Sanchez & Schmitt (2010:36-39) tested their participants’ vocabulary acquisition through personal interviews alone, during which each participant explained in their own words what they learned about the target words.

In this study, the In-Depth Interviews were used to try to ascertain the possible factors that may explain any vocabulary acquisition revealed in the vocabulary tests (see Section-3.4 p.72 & Section-4.2 p.135). Hence the quantitative data (i.e. the vocabulary tests, reading diaries & surveys) was principally for giving insights into what the participants were like, while the qualitative data (i.e. the Supplementary Questioning & In-Depth Interviews) was chiefly for why they were like that, remembering that even though the quantitative data was also meticulous and comprehensive and could be used to deduce some of these reasons, the qualitative data would reveal more directly from what the students said. This methodology backs up quantitative data with qualitative data, which is often done by other researchers (Benson, 2001:196) because using different perspectives, methods and sources of information can greatly improve the authority of the ensuing results (Chappelle & Duff, 2003:165).

Interviews are particularly useful for adding a qualitative dimension to a study, because they allow participants to ‘express how they see situations from their own perspective’ (Spears, 1995:181), and their use recognizes ‘that people themselves are a valuable

source of information' (Lederman, 1990:118) and that human interaction is central to understanding the social context of research data (Hyland, 2002:181). Hence, in the field of vocabulary acquisition, Li & Schmitt (2009:87-90) had a quantitative numerical score assigned to the degree of appropriateness of each lexical phrase found in their participant's writing, but they then interviewed the participant to qualitatively ascertain her confidence in using those phrases. Furthermore in the field of extensive reading, Macalister (2008:251-254) used quantitative questionnaires with qualitative interviews to investigate (amongst other things) the participants' attitudes to his ER programme.

Once the In-Depth Interviews of the Main Study were completed, the enormous quantity of information contained therein needed a precise and consistent procedure of analysis. This is detailed in full at the start of the Results section for the In-Depth Interviews (see Section-4.7 p.164).

3.9.1 Participants

The participants of the In-Depth Interviews were carefully chosen to be as representative as possible of the Experimental Cohort, as advised by Nunan (1992:152) even for small-scale studies. A final total of twelve students were selected from the Main Study, which was a considerable number given that they represented almost 25 per cent of the Experimental Cohort (n=49), that they were chosen to be as representative as possible, and that their interviews were carefully planned, long, in-depth and detailed, as expounded in this section.

A systematic process was adopted to select these participants, using data from the Vocabulary Tests (Section-4.2 p.135), End-of-Semester Exams (Section-4.3 p.149) and the Survey of Exposure to English Learning (Section-4.5 p.156). Hence, firstly, they were chosen to have a mix of language abilities, in terms of vocabulary test scores, and End-of-Semester Exam results. They were also chosen to be a representative mix in terms of age and life experience, and so most (ten out of the twelve) were young bachelors, while a few (two out of the twelve) were older married men. These were considered relevant because age is an important psychological factor in language

learning (Benson, 2004:10), which could be related to background knowledge, which is thought to aid text comprehension (Nation & Coady, 1988:102-104; Schmitt et al., 2011:38). Also, marital status could be an important factor, given that the researcher and most of his colleagues have informally observed that the married students here have a more mature and responsible attitude to studying than the single young men. Finally, they were chosen to be from a representative selection of towns and cities within the kingdom, noting that in the view of the researcher who has spent many years working in this context, there is a large variation of cultures in the various towns of this vast country that could affect their learning attitudes and abilities. Hence 2 were from the capital, 1 from the large city in the western region, 3 from large towns in the western region, 3 from small towns in the central region and 3 from small towns in the southern region.

The quantitative aspects of their profile were precisely compared with the corresponding aspects of the entire Experimental Cohort, to ensure there were no statistically significant differences. Age was taken from the Survey of Exposure to English Learning (Section-4.5 p.156), End-of-Semester Exam results were taken from Section-4.3 p.149 and Vocabulary Test scores were taken from Section-4.2 p.135 (only using those results that yielded statistically significant vocabulary gains). This analysis is summarised in Appendix-16, showing that out of the 38 separate analyses, all except just one were not significantly different ($p < 0.05$ for significant difference using the Mann-Whitney U Test for non-normally distributed independent samples – see Pallant, 2010:227-230). For example the In-Depth Interview Participants' results for the End-of-Semester Exam (Median=88%, $n=12$) were not significantly different to those of the entire Experimental Cohort (Median=83%, $n=49$), $U=227$, $z=-1.22$, $p=0.224$. This clearly demonstrated that for these characteristics, the participants of the In-Depth Interviews were an acceptable representation of the whole Experimental Cohort.

3.9.2 Procedure

The procedure for the In-Depth Interviews was carefully planned and administered to ensure that the data yielded would be as reliable as possible and as unbiased as possible.

One fundamental potential problem is that the participants being interviewed may not give their true opinions. For example, they may simply want to please the interviewer, and so describe the ER programme positively, feeling that this is what the interviewer wanted to hear. Also, the special attention being given to them during the interview could artificially generate a greater quantity of feedback that could also be more positive than is actually the case. To mitigate this, the interviews were firstly all held after the final exams of the preparatory course (in which the ER programme and vocabulary test were incorporated), and so the interviewer (the researcher) was no longer their teacher and no longer had any role in the participants' college grades.

The interviews were also held at times chosen by the participants, to ensure they were as relaxed as possible, and hence able to be as frank and complete as possible in what they said. Both of these points were explicitly communicated to the interviewees at the start of the interview, and the introduction of the interviews also included a specific request that their views should be complete and honest to enable the researcher to use them effectively in his research and his future teaching, and that without this there would be no benefit. In addition, after asking the interviewees for their own opinions, they were sometimes also asked about the opinions they knew of the rest of class. This was useful to gauge the wider opinions of the rest of the participants in the study, but it also allowed the interviewees to give negative feedback that they may feel more comfortable voicing as the opinion of others. More about this is given below when detailing the Interview Schedule (Section-3.9.2.1 p.124).

A conscious decision was made to only use audio recordings, agreeing with Dörnyei (2007:139) who considered video recording equipment to be much more obtrusive than audio recording equipment. It was also decided that all interviews would be transcribed, following the general consensus reported by Dörnyei (*ibid.*, p.139) that writing notes can lead to the loss of important details, and can lead to the interview process being disrupted. Nunan (1992:153) also added that although transcribing recordings can be time-consuming and lead to data overload, they are nonetheless naturalistic, objective and can be analysed again and again. Although note-taking mitigates the need for time-

consuming transcriptions, there exists the fundamental disadvantage of recorder bias, which may lead to the status of the whole data being questioned. A complete transcript of one of the interviews is given in Appendix-22, but taking the view of Braun & Clarke (2006) that the actual transcription process was itself an important stage during analysis of the results, the methodology of the transcription is detailed later in the Results Section-4.7 p.164.

The first interview was what Lederman (1990:117) describes as a ‘focus group interview’, which was used to pilot procedures and generate preliminary ideas. Four participants were chosen, because this number was thought to be large enough to encourage the emergence of what Dörnyei (2007:144) describes as a ‘collective wisdom’, and yet small enough to allow everyone to participate. These four were selected to be of very different language abilities, and they were also selected because they were known to get on well with each other, in order to have a relaxed and friendly atmosphere that would encourage many of the benefits Lederman (1990:119-120) noted for focus group interviews, such as gaining more information in a shorter time, generating an overall voice in addition to the individual voices of the participants, and making the process more of a multi-directional discussion than a just a two-way dialogue.

It could be pointed out that such a group interview may lead to the more vocal participants overshadowing the quieter and weaker ones (Dörnyei, 2007:146), but the interviewer (i.e. the researcher), being a practising teacher, was able to actively bring out responses from all the participants in a similar way he would with the quieter and weaker students in class during a normal lesson. Following the advice of Dörnyei (*ibid.*, pp.145-146), the researcher’s role during the focus group interview was more to facilitate discussion and probe emerging themes, than to interview in the traditional sense.

This first focus group interview yielded results that were directly relevant to RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8

p.36), and its whole process gave ideas that moulded the administration of future individual interviews.

For example, the one hour duration seemed to be an ideal length in terms of providing the opportunity to yield in-depth information, with this duration being the upper limit of what is reported to be typical in qualitative interviews (Dörnyei, 2007:134), and similar to that used in previous vocabulary acquisition studies such as Pellicer-Sanchez & Schmitt (2010:40). However, it was clear that the weaker English speakers found it difficult to express themselves for this long time without breaking into their local dialect of Arabic, which the researcher had difficulty understanding. Hence the subsequent individual interviews were held in one session for the more fluent English speakers, but were split into two parts for the weaker speakers. This splitting of their time also gave a useful chance to ponder over the participants' responses during the first half, which then led to more precise follow-up prompts and questions being formulated and posed during the second half. This follows the recommendations of some researchers, as reported by Dörnyei (2007:134-135), who specifically endorse this kind of 'sequence' of interviews to gain greater depth and focus, and to allow both the interviewer and interviewee to think more intensely about the experience. Interviewing participants with weak English during the focus group interview also led to subsequent questions and phrases being refined and simplified, as detailed below in Section-3.9.2.1 p.124 when describing the Interview Schedule.

The subsequent individual interviews were administered with a systematic and consistent procedure that was rigorously planned, but balanced with a large degree of flexibility. Semi-structured interviews were used to allow for this planned but flexible procedure, with the same Interview Schedule (see Section-3.9.2.1 below) being referred to in order to safeguard consistency as much as possible. This complemented the heavily structured Reading Diaries (Section-3.6 p.103) and Surveys (Section-3.7 p.111) given before, and which followed the view reported by Li & Schmitt (2009:88) that interviews should be 'guided conversations rather than structured queries'. Semi-structured interviews have prepared guiding questions and prompts in an open-ended format that encourages the interviewee to elaborate on issues raised (Dörnyei,

2007:136), and they are popular with researchers (Nunan, 1992:149) because they give the interviewer guidance, direction (Dörnyei, 2007:136) and flexibility (Nunan, 1992:150), and because they also offer the interviewee more control of the course of the interview (*ibid.*, p.150).

As mentioned above in Section-3.9 p.118, once the procedure for all the Main Study In-Depth Interviews was completed, the large amount of data that resulted needed to undergo a carefully thought-out process of analysis. This methodology of analysis is detailed in full at the start of the Results section for the In-Depth Interviews (see Section-4.7 p.164).

3.9.2.1 *Interview Schedule*

The Interview Schedule is given in Appendix-17, and shows that questions were worded to be as simple and as easy to understand as possible. The first part was what Nunan (1992:152) termed the ‘briefing’, in which the interviewee was thanked for their participation, briefed about the purpose of the interview, and told that it was to be recorded. Then came the introduction and initial questions to ‘set the tone and create initial rapport’ (Dörnyei 2007:137), making the interviewees relaxed, and impressing that their views were important (*ibid.*, p.140). Hence, they were asked to be frank as possible, being reminded that the interviewer was no longer their teacher and that he no longer had any role in their college grades. They were also asked to be as detailed as possible, being reminded that the timing of the interview was their choice so that they would not feel rushed or pressured. Hence they were explicitly told that without their honest and complete feedback, the interviews would be of little use to the interviewer’s research or future teaching.

The next part of the Interview Schedule was what Dörnyei (2007:137) described as ‘content questions’, aimed at inviting the interviewees to explain their own opinions about possible factors that may explain any vocabulary acquisition revealed in the vocabulary tests (see Section-3.4 p.72 & Section-4.2 p.135), hence addressing RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8

p.36). Leading questions were avoided as much as possible (as advised by Dörnyei, 2007:138), and instead the questions were carefully worded to be as open as possible, to allow the interview to be ‘guided by the responses of the interviewee rather than the agenda of the researcher, enabling unanticipated themes and topics to emerge’ (Spears, 1995:181). They were based on potential factors suggested by the other collected data, aiming to examine the psychology, educational and social situation of the students that could explain their reading habits and vocabulary gains. As recommended by Dörnyei (2007:137-138), they focused on the interviewees’ experiences, behaviours, opinions, values, feelings, what they reported to know, and what they described about their social and demographic backgrounds.

Each of the content questions described above, consisted of ‘probes’ (Dörnyei, 2007:138) designed to elicit richness and depth in the initial responses by being more specific. Some of these were core probes, to ensure consistency in approach, while others were possible follow-up probes that allowed for divergent topics and branches of discussion. Many possible follow-up probes were planned because many of the interviewees were not so fluent in their spoken English, and so they found it difficult to continue without being prompted. As also seen in the Interview Schedule (Appendix-17), the interviewees were sometimes asked about the opinions of the rest of class. This was to indirectly elicit the views of the other participants of the experimental cohort, but it also gave the interviewees the option of giving negative feedback that they may feel more comfortable voicing as the opinion of others.

The last part of the Interview Schedule (Appendix-17) was the ‘final closing’ (Dörnyei, 2007:138) which invited the interviewees to have the final say, offering them to add anything else they wished. Also during this final stage, sincere thanks were offered once again for the interviewees’ feedback, as advised by Dörnyei (*ibid.*, p.143).

The style used to execute the Interview Schedule reflected the key goal of attempting to elicit as much information from the participants themselves to investigate the possible factors that may explain any vocabulary acquisition revealed in the vocabulary tests (see Section-3.4 p.72 & Section-4.2 p.135), and so address RQ2 (concerning the impact of

other factors on vocabulary acquisition) (see Section-2.8 p.36). Firstly, as advised by Dörnyei (2007:138), simple language was consciously chosen to make it easy for the interviewees to understand and hence respond appropriately, and Arabic was sometimes used when needed. Secondly, the interviewees were (as much as possible) simply left to speak, agreeing with Dörnyei's view (*ibid.*, p.140) that the interviewer's primary job is simply to listen, and this was relatively easy with the more fluent speakers. However, even when weaker speakers seemed to 'break-down', running out of words to express themselves, the initial interviewer's response was to simply try to remain silent, resisting the urge to step in too quickly, with these 'silent probes' (*ibid.*, p.142) being used as a cue for more information. Only when this did not encourage more information, were other techniques used to encourage elaboration (see Dörnyei, 2007:142), such as 'echo prompts' for which the last words of the interviewee were repeated, and reflective summary for which the last point made by the interviewee was summarized. In order to maintain the flow of information, simple cues of feedback were continuously given, similar to those advised by Dörnyei (*ibid.*, p.142) such as 'carry-on feedback' (gestures like nods, words like 'yes', and affirmative sounds like 'a-ha'); 'reinforcement feedback' (exclamations of appreciation like 'that's excellent – I never thought of that'); and even 'negative reinforcement' that is needed when responses have strayed off topic (invitations like 'please let me go back to what you said before so I can understand you fully').

Hence the style described above for executing the Interview Schedule was to allow the interview to flow naturally, and it was also to encourage the interviewees to give as much rich detail as possible, which Dörnyei (2007:140) considers are the two key features of a good qualitative interview. As a result, the Interview Schedule was actually more of an 'interview guide' (*ibid.*, pp.136-137), which simply provided the framework of broad questions, and which did not seek to limit the depth and breadth of interviewee responses, or inhibit unstructured responses that may deviate from those anticipated by the interviewer.

3.9.3 Compensating for Potential Limitations

As described above in Section-3.9.2 p.120 and Section-3.9.2.1 p.124, extensive efforts were made to compensate for the potential limitations that are inherent in any interview process, to ensure that the methodology was robust enough to yield results that were as reliable and as unbiased as possible.

Hence in summary of what was described above, twelve participants were interviewed, which was considerable given that this represented almost 25 per cent of the experimental cohort (n=49), and given that the interviews were long, carefully planned, in-depth and detailed. In addition, the participants were carefully chosen through a systematic process to ensure they were as representative as possible of the whole Experimental Cohort.

Also, long interviews generate lots of detailed information that can be difficult to analyse, and so to address this, all of them were recorded and transcribed. This process was deemed to be more naturalistic, more objective, more complete and less obtrusive than writing notes, and also offered opportunities to easily revisit the data at any time. In addition, the social and spontaneous nature of interviews can make comparisons between participant responses difficult. To address these issues, the researcher adopted the same systematic and consistent approach that was administered through the same clearly planned interview schedule, making like-for-like comparisons easier.

During the first focus group interview, the more confident speakers could have dominated with their opinions, but this was continuously countered with active efforts to give all participants as many opportunities as possible to contribute. For the subsequent individual interviews, weaker speakers could have found it difficult to express themselves clearly during the long interview, and so to mitigate this problem, their interviews were split into two shorter sessions.

Furthermore, there could be a fear that the participants may give limited, narrow and shallow responses, and so a conscious decision was made to use semi-structured

interviews with prepared guiding questions and probes in an open-ended format that encouraged the interviewees to speak from their inner selves and to elaborate on issues they raised, while the interviewer made constant efforts to simply listen and let the interviewees talk as much as possible.

Similarly, the interviewees may not give their true opinions, or may exaggerate out of the excitement of being the centre of attention, or they may report what they felt the interviewer wanted to hear in order to make him happy. All of this was countered with specific measures, such as holding the interviews at times chosen by the interviewees (to enable them to be as relaxed as possible), and specifically reminding them that the interviewer was no longer their teacher and no longer had influence over their college grades. They were also explicitly reminded that without complete and honest responses, the whole interview would not be useful for the current research or the planning of future of courses.

Finally, there were problems that could occur during the interview itself that needed careful attention and planning to mitigate. One very real problem was that the weaker speakers may not understand the questions, and so careful planning went into their wordings, and Arabic translations were given when needed. Also, it was quite possible that interviewees would go off the topic, and so polite phrases of negative reinforcement were pre-prepared and used to mitigate this. In addition, it was likely that participants would pause or stop out of lack of confidence or out of not being sure how detailed their answers should be. This too was mitigated with pre-planned phrases, but this time containing carry-on feedback and reinforcement feedback.

All of these measures were taken to try to mitigate or compensate for the limitations and problems faced during the interview process, making the methodology more robust, and hence ensuring the results yielded were as reliable as possible.

3.10 METHODOLOGY FOR THE RESEARCH JOURNAL

Throughout the study, the researcher kept a Research Journal, noting down as much as possible that seemed to be of relevance at the time. This consisted of many field notes, which were simple observations, thoughts, reactions and questions that arose during the different events and stages of the project, and hence they were not of any one precise form, taking the view of Richards (2003:135-136) that field notes are no more than a personal subjective spontaneous account of a researcher's encounter with the world being investigated. With this in mind, the notes took all the forms described by Richards (*ibid.*, p.137), namely mental notes, jotted notes and full notes, and they were often initially recorded in different locations. For example, some of the field notes were made on a dedicated first sheet of each of the many spreadsheet books the researcher made to record different aspects of the project, especially those related to data collection. Other notes relating to discussions and meetings with, for example the supervisor, were kept in word processing files, while hand written notes were made directly onto student specific material when questioning them about interesting or unusual entries. In addition, photographs were taken in situations when nothing but the researcher's phone was at hand. In all these cases, the note taking was immediate to preserve their accuracy of content, as recommended by Richards (*ibid.*, p.136), Benson (2004:13-14) and others. The notes that were entered electronically into spreadsheets and word processing documents were easily and accurately copied and pasted during the collating stage, and photos of non-digital notes were kept electronically and filed in the relevant folders set up by the researcher. However, the photos (even those containing just handwritten notes) were usually not fully re-written, out of fear that some of the original connotations would be lost during the re-writing process.

This large body of varied field notes helped the researcher to organize every stage of the study, enabling him to 'notice interesting events' and formalize them into 'critical events' (as described by McDonough & McDonough, 1997:80-82) that were key in moulding the direction of the research. The Research Journal also enabled the researcher to reflect on his findings at different stages of the study, as advised by Richards (2003:114) who recommended that such reflection should be at different

stages to generate different insights from different vantage points. It also allowed for patterns to be spotted (Hyland, 2002:188), that then formed the bases of further lines of enquiry, for example, when determining the main questions that would form the core of the In-Depth Interviews that followed the testing, diaries and surveys.

Finally, the Journal also gave the opportunity for the researcher to embark on an on-going 'critical reflection' (as termed by Farrell, 2001:24) of his many roles in the project – teacher, researcher and post-graduate student – to improve himself through constructive self-criticism of all aspects of his professional, educational and social activities during the project, with a journal being particularly useful for this because it is a tool that is both introspective and retrospective (Hyland, 2002:188).

CHAPTER-4

RESULTS

In the previous chapter, a detailed description of the Methodology was given. In this chapter, a detailed description of the Results is now presented.

As stated before in Section-2.8 p.36 this study aimed to address the following research questions:

Research Question 1 What is the impact of an extensive reading (ER)
(RQ1) programme of a relatively short duration, on the
 vocabulary acquisition of male Saudi university students?

Research Question 2 What characteristics of the students, in terms of the nature
(RQ2) of their reading, their other English learning, and their
 culture, may also be associated with any observed
 vocabulary acquisition?

In order to address RQ1, pretests and posttests of vocabulary knowledge were administered to the experimental and control groups to gauge vocabulary gain. In order to address RQ2, reading diaries were completed, a survey of other sources of English learning was administered, supplementary questioning were posed, and in-depth interviews were conducted, all to gauge factors that could be associated with any vocabulary gain. Some of this data was more quantitative in nature, while other data was more qualitative, taking a view that both are complementary and equally important (Burns, 1999:24), and taking a further view that the distinction between the two is not always clearly defined (Nunan, 1992:3; Richards, 2003:11).

The development of these data collection instruments took place in 3 stages, as detailed earlier (see Section-3.1.1 p.42), and as summarised again below in Table-2 (duplicated):

Table-2 (duplicated): A Summary of the Stages of this Study

STAGE	DESCRIPTION	PARTICIPANTS	ROLE
Stage-1	Initial Pilot	Expert focus group (n=7) of English Language teachers	Piloting the initial versions of the vocabulary tests
		Student focus group (n=4)	Piloting the initial data collection instruments & ER Programme
Stage-2	Main Pilot	Large pilot cohort of students (n=52)	Piloting the revised data collection instruments & ER Programme
Stage-3	Main Study	Experimental cohort (n=49) and Control cohort (n=36)	Using the finalised data collection instruments & ER Programme

Hence during the Main Study, the final data collection instruments were as summarised again below in Table-4 (duplicated):

Table-4 (duplicated): Data Collection Instruments during the Main Study

	DESCRIPTION	QUANTITY	RELEVANT RQ	DETAILED IN:
1	Institution English Placement Tests	n=85 scripts	Sampling & Class Allocation for RQ1 & RQ2	Section-3.3 p.69 & Section-4.1 p.134
2	Pretests and Posttests developed by the researcher, to gauge different aspects of Vocabulary Acquisition	n=680 scripts for 4 separate tests, gauging 3 different aspects of partial vocabulary knowledge, and testing 200 words	RQ1	Section-3.4 p.72 & Section-4.2 p.135
3	Institution End-of-Semester English Exams	n=85 scripts	Check effect of ER Programme on Participants' exam results	Section-3.5 p.101 & Section-4.3 p.149
4	Reading Diaries developed by the researcher to record student reading patterns	n=49 diaries, recording an aggregate of approximately 600 weeks of reading	RQ2	Section-3.6 p.103 & Section-4.4 p.151
5	Surveys developed by the researcher to record Other Sources of Exposure to English	n=49 surveys	RQ2	Section-3.7 p.111 & Section-4.5 p.156
6	Supplementary Questioning	hundreds over a period of 2 years	RQ1 & RQ2	Section-3.8 p.117
7	In-Depth Interviews	12 participants (approx. 25% of the Experimental Cohort), for approximately one hour each	RQ2	Section-3.9 p.118 & Section-4.7 p.164
8	A Research Journal containing notes, observations and events	spanning the duration of the project	RQ1 & RQ2	Section-3.10 p.129

Chapter-3 (Methodology) gave feedback from the Initial Pilot, and preliminary results from the Main Pilot.

What follows now is a detailed description of the analysis and results for each of these data collection instruments during the Main Study.

4.1 PLACEMENT TEST RESULTS

As detailed in the Methodology Chapter-3 (Section-3.3 p.69), participants from the Main Study (n=85) (see Section-3.1.1 p.42) completed the institution's English Language placement test during their first lesson, and this was used for class allocation as a preliminary stage to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36). Details of the test, its administration and its use in assigning students of similar abilities to the experimental and control cohorts are also given in Section-3.3 p.69.

Excel spreadsheets and IBM's SPSS (v.17 and higher) were employed to process the data to give results for both the descriptive and inferential statistics (see Appendix-14). The descriptive statistics (Pallant, 2010:59-63) showed that the data was not normally distributed (see the histograms in Appendix-14), and so the non-parametric Mann-Whitney U Test for 2 independent samples was used (*ibid.*, p.227), where $p < 0.05$ indicated significant difference between the groups (*ibid.*, pp.240-2).

4.1.1 Results

The Mann-Whitney U Test (see Appendix-14) showed there was no significant difference between the Placement Test results of the Main Study Experimental cohort (Median=45%, n=45) and the Main Study Control cohort (Median=33%, n=32), $U=549$, $z=-1.77$, $p=0.077$. This gave confidence that it would be valid to compare the two after treatment, and that any differences arising would not be those that existed before the treatment.

4.2 VOCABULARY TESTS RESULTS

As detailed in Chapter-3 (Methodology), in order to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36), four vocabulary tests were developed during the Initial Pilot and the Main Pilot, and the final versions were administered as pretests and posttests to the Experimental and Control cohorts of the Main Study. These four tests were detailed in Section-3.4.3 p.76, and are briefly summarised below in Table-12:

Table-12: Brief Summary of the Vocabulary Tests Used in This Study

	Test-1	Test-2	Test-3	Test-4
Word Knowledge tested	Receptive Written Form i.e. SPELLING	Productive Form-Meaning i.e. OPEN TRANSLATION	Receptive Form-Meaning i.e. MEANING	Receptive Form-Meaning i.e. MEANING
Kind of Test	Multiple Choice	Open Response	Multiple Choice	Multiple Choice
Source of Target Words	Vocabulary Size Test (Nation, 2009)	Vocabulary Size Test (Nation, 2009)	Vocabulary Size Test (Nation, 2009)	Vocabulary Levels Test (Schmitt et al., 2001)
Total Number of Target Words tested	n=50	n=50	n=50	n=150
Word Frequency Level Groupings	1000-5000 & 6000-10000	1000-5000 & 6000-10000	1000-5000 & 6000-10000	2000, 3000, Academic, 5000 & 10000
Target Words per Grouping	25 & 25	25 & 25	25 & 25	30, 30, 30, 30 & 30

The Methodology Chapter-3 (Section-3.4.2 p.75 to Section-3.4.6 p.89) precisely detailed how the tests were devised and administered to enable partial vocabulary acquisition to be determined in three different ways, namely: 1. testing different facets of word knowledge; 2. testing vocabulary from different levels of word frequency; 3. accounting for different levels of certainty for any given answer. This resulted in a large amount of data that called for a cautious and constant methodology of analysis to safeguard the internal reliability of the research, whereby other independent researchers could re-analysis the same data to produce comparable results (Nunan, 1992:14-17). This methodology was fully expounded in Section-3.4.7 p.90 for the Main Pilot group and this same methodology was used for the experimental and control cohorts of the Main Study.

Hence Section-3.4.7 p.90 meticulously explained how the results were analysed separately for each of the 4 tests, and then how each test's results were processed independently according to the different word frequency levels of the vocabulary items, and then how each of these results were analysed separately for different levels of certainty the participants recorded for each test item entry.

Thus the analysis was separated into:

1. 4 separate tests, gauging 3 different aspects of word knowledge;
2. Each of the above contained a pretest and posttest;
3. Test-1 (Spelling), Test-2 (Open Translation) and Test-3 (Meaning) (all based on Nation, 2009) tested 50 words that were separated into 2 groups of word frequency;
4. Test-4 (Meaning) (based on Schmitt et al., 2001) tested 150 words that were separated into 5 groups of word frequency;
5. All of these were then further separated into Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone.

Therefore, a total of 66 sets of data were separately analysed, for each of the 52 participants of the Main Pilot, for all of the 200 word items of vocabulary, as shown in Figure-6, Figure-7 & Figure-8 below, and as shown more in Appendix-18.

Like in the analysis of the Main Pilot data (see Section-3.4.7 p.90), a large number of Excel spreadsheets were used to manipulate this vast quantity of data, and IBM's SPSS software (v.17 and higher) was used to process it, to yield results for both the descriptive and inferential statistics. The descriptive statistics included means, standard deviations, medians, and quartiles, while the inferential statistics included Cronbach's alpha coefficient for mean inter-item correlation to check the internal consistency (Pallant, 2010:97) of the tests, and also included hypothesis testing to compare the pretests and posttests for significant gain. As seen below in Section-4.2.5 p.144, the descriptive statistics showed that the data was not normally distributed, and so the non-parametric Wilcoxon Signed-Rank Test for repeated measures was used to compare the pretests and posttests (*ibid.*, p.230).

The results of the vocabulary tests during the Main Study are given in Appendix-18, comparing the Experimental and Control cohorts to address RQ1 (concerning the impact of the ER programme on vocabulary acquisition) (see Section-2.8 p.36). These are summarised below in Figure-6, Figure-7 & Figure-8, and are further detailed in the following Section-4.2.1 p.141 to Section-4.2.7 p.147.

Figure-6: Summary of Vocabulary Test Results during the Main Study comparing Experimental & Control Cohorts (change in medians given to 1 d.p.) [See Appendix-18]

TEST-1 Based on NATION (2009)

SPELLING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	10.0	10.4	3.6	10.0	10.1	3.9	11.0	11.8	3.4	11.0	11.0	4.7	1.0	1.0
	Correctness x Certainty	7.7	8.0	3.2	8.0	7.9	3.5	9.3	9.8	3.4	8.5	9.2	4.4	1.7	0.5
	Certainty only	12.3	12.0	4.1	13.0	12.1	4.1	14.3	14.3	3.7	13.0	13.5	4.7	2.0	0.0
6000-10000wd 25 items	Correctness only	5.0	5.4	3.5	4.5	5.2	3.6	5.0	6.4	4.6	5.0	5.3	3.7	0.0	0.5
	Correctness x Certainty	2.7	3.3	2.2	2.3	3.2	2.6	3.3	4.2	3.4	2.8	3.7	3.2	0.7	0.5
	Certainty only	6.0	7.2	4.4	7.0	7.8	4.9	8.0	8.1	4.8	7.8	7.9	5.1	2.0	0.8

TEST-2 Based on NATION (2009)

TRANSLATION

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	7.0	6.8	2.8	5.5	6.6	3.6	9.0	8.7	3.1	6.5	7.4	3.7	2.0	1.0
	Correctness x Certainty	6.0	6.3	2.7	5.2	6.1	3.4	8.3	8.0	3.0	6.0	7.0	3.8	2.3	0.8
	Certainty only	8.7	9.0	3.2	8.2	8.4	3.9	10.7	11.0	3.4	9.0	9.3	4.4	2.0	0.8
6000-10000wd 25 items	Correctness only	1.0	0.8	0.9	1.0	1.0	1.9	1.0	1.4	1.4	1.0	1.6	2.1	0.0	0.0
	Correctness x Certainty	0.3	0.6	0.7	0.5	0.8	1.7	1.0	1.1	1.2	0.8	1.3	2.0	0.7	0.3
	Certainty only	2.3	2.7	2.1	2.0	2.7	3.1	2.7	3.4	2.5	2.0	3.1	3.4	0.3	0.0

TEST-3 Based on NATION (2009)

MEANING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	9.0	9.4	3.7	8.0	8.7	4.4	10.0	10.6	3.5	9.0	9.0	4.4	1.0	1.0
	Correctness x Certainty	7.3	7.7	3.3	6.3	7.0	4.1	8.3	8.8	3.3	6.7	7.7	4.3	1.0	0.3
	Certainty only	13.0	12.5	3.6	11.0	12.4	4.7	14.3	13.9	4.0	14.3	13.6	4.6	1.3	3.3
6000-10000wd 25 items	Correctness only	4.0	3.8	2.1	3.0	4.2	3.4	4.0	4.5	3.0	3.0	4.1	2.7	0.0	0.0
	Correctness x Certainty	2.0	2.1	1.3	2.0	2.6	2.5	2.3	2.7	2.0	2.0	2.7	2.2	0.3	0.0
	Certainty only	6.0	6.4	3.3	7.0	7.3	5.1	6.3	7.1	4.2	7.0	7.9	4.7	0.3	0.0

TEST-4 Based on SCHMITT ET AL (2001)

MEANING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
2000wd 30 items	Correctness only	7.0	9.0	5.7	7.0	8.6	6.5	11.0	11.3	6.2	8.5	10.1	6.9	4.0	1.5
	Correctness x Certainty	5.0	6.6	5.1	3.7	6.3	6.0	7.0	8.9	6.0	6.2	8.1	6.9	2.0	2.5
	Certainty only	13.3	13.0	5.4	11.8	13.4	6.7	15.7	15.6	5.7	16.5	16.0	6.7	2.3	4.7
3000wd 30 items	Correctness only	5.0	6.5	4.6	4.5	6.2	5.6	7.0	7.8	5.4	6.0	7.8	6.0	2.0	1.5
	Correctness x Certainty	3.7	4.6	4.1	3.0	4.6	5.0	4.3	5.9	4.9	3.5	5.7	5.9	0.7	0.5
	Certainty only	8.7	9.7	5.5	10.0	11.3	7.4	10.7	11.4	5.8	13.5	12.4	6.7	2.0	3.5
Academic wd 30 items	Correctness only	6.0	7.1	5.3	5.0	6.7	6.0	7.0	8.2	5.7	7.0	8.4	6.7	1.0	2.0
	Correctness x Certainty	3.0	5.1	4.7	3.3	5.1	5.5	4.3	6.4	5.6	4.3	6.4	6.3	1.3	1.0
	Certainty only	8.7	9.9	5.7	11.2	11.4	6.6	12.0	12.6	6.7	13.3	12.5	7.2	3.3	2.2
5000wd 30 items	Correctness only	3.0	3.4	2.9	3.0	3.6	4.5	3.0	4.0	3.2	4.0	4.8	5.5	0.0	1.0
	Correctness x Certainty	1.3	2.0	1.9	1.5	2.3	3.8	2.0	2.4	2.4	2.0	3.1	4.7	0.7	0.5
	Certainty only	6.0	6.6	4.6	6.7	7.9	5.7	8.0	8.2	5.3	9.5	9.2	6.1	2.0	2.8
10000wd 30 items	Correctness only	1.0	1.8	2.2	1.0	2.2	2.6	1.0	2.0	2.4	2.0	3.2	3.9	0.0	1.0
	Correctness x Certainty	0.7	1.0	1.3	0.7	1.1	1.8	0.3	1.0	1.5	0.8	1.7	2.6	-0.3	0.2
	Certainty only	4.0	4.7	4.4	4.8	5.4	5.1	4.0	5.3	4.8	6.0	6.9	6.0	0.0	1.2

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." Language Testing 18 (1), pp.55-88.

*Significant Change is in GREEN fill; Non-Significant Change is in RED text.

Figure-7: Summary of Vocabulary Test Results during the Main Study - Experimental Cohort [See Appendix-18]

TEST-1 Based on NATION (2009)
SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.670	10.0	0.647	11.0	0.000	-3.49	0.35	1.0	10%
	Correctness x Certainty	0.751	7.7	0.779	9.3	0.000	-4.61	0.47	1.7	22%
	Certainty only	0.882	12.3	0.859	14.3	0.000	-4.54	0.46	2.0	16%
6000-10000wd 25 items	Correctness only	0.731	5.0	0.841	5.0	0.050	-1.96	0.20	0.0	0%
	Correctness x Certainty	0.684	2.7	0.853	3.3	0.023	-2.27	0.23	0.7	25%
	Certainty only	0.904	6.0	0.919	8.0	0.156	-1.42	0.14	2.0	33%

TEST-2 Based on NATION (2009)
TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.717	7.0	0.736	9.0	0.000	-5.00	0.50	2.0	29%
	Correctness x Certainty	0.750	6.0	0.763	8.3	0.000	-5.08	0.51	2.3	39%
	Certainty only	0.773	8.7	0.792	10.7	0.000	-4.57	0.46	2.0	23%
6000-10000wd 25 items	Correctness only	0.194	1.0	0.423	1.0	0.003	-2.95	0.30	0.0	0%
	Correctness x Certainty	0.182	0.3	0.434	1.0	0.001	-3.41	0.34	0.7	200%
	Certainty only	0.674	2.3	0.721	2.7	0.028	-2.20	0.22	0.3	14%

TEST-3 Based on NATION (2009)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.722	9.0	0.687	10.0	0.003	-3.01	0.30	1.0	11%
	Correctness x Certainty	0.772	7.3	0.774	8.3	0.000	-3.73	0.38	1.0	14%
	Certainty only	0.843	13.0	0.893	14.3	0.001	-3.41	0.34	1.3	10%
6000-10000wd 25 items	Correctness only	0.290	4.0	0.633	4.0	0.091	-1.69	0.17	0.0	0%
	Correctness x Certainty	0.301	2.0	0.662	2.3	0.099	-1.65	0.17	0.3	17%
	Certainty only	0.838	6.0	0.909	6.3	0.473	-0.72	0.07	0.3	6%

TEST-4 Based on SCHMITT ET AL (2001)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.867	7.0	0.877	11.0	0.000	-4.39	0.44	4.0	57%
	Correctness x Certainty	0.907	5.0	0.918	7.0	0.000	-4.93	0.50	2.0	40%
	Certainty only	0.908	13.3	0.923	15.7	0.000	-4.25	0.43	2.3	18%
3000wd 30 items	Correctness only	0.817	5.0	0.860	7.0	0.005	-2.80	0.28	2.0	40%
	Correctness x Certainty	0.871	3.7	0.901	4.3	0.000	-3.88	0.39	0.7	18%
	Certainty only	0.920	8.7	0.928	10.7	0.012	-2.52	0.25	2.0	23%
Academic wd 30 items	Correctness only	0.867	6.0	0.878	7.0	0.029	-2.18	0.22	1.0	17%
	Correctness x Certainty	0.906	3.0	0.920	4.3	0.000	-3.54	0.36	1.3	44%
	Certainty only	0.927	8.7	0.949	12.0	0.000	-3.62	0.37	3.3	38%
5000wd 30 items	Correctness only	0.671	3.0	0.711	3.0	0.254	-1.14	0.12	0.0	0%
	Correctness x Certainty	0.716	1.3	0.780	2.0	0.063	-1.86	0.19	0.7	50%
	Certainty only	0.909	6.0	0.934	8.0	0.047	-1.99	0.20	2.0	33%
10000wd 30 items	Correctness only	0.667	1.0	0.703	1.0	0.697	-0.39	0.04	0.0	0%
	Correctness x Certainty	0.674	0.7	0.759	0.3	0.774	-0.29	0.03	-0.3	-50%
	Certainty only	0.932	4.0	0.950	4.0	0.662	-0.44	0.04	0.0	0%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill. Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill. Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{N(\text{pretest} + N\text{posttest})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = ((posttest score – pretest score) / pretest score) x 100. Note if pretest score = 0.0, then Percentage Change cannot be calculated.

Figure-8: Summary of Vocabulary Test Results during the Main Study - Control Cohort
[See Appendix-18]

TEST-1 Based on NATION (2009)
SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.718	10.0	0.826	11.0	0.181	-1.34	0.17	1.0	10%
	Correctness x Certainty	0.784	8.0	0.868	8.5	0.060	-1.88	0.24	0.5	6%
	Certainty only	0.878	13.0	0.909	13.0	0.049	-1.97	0.25	0.0	0%
6000-10000wd 25 items	Correctness only	0.747	4.5	0.765	5.0	0.898	-0.13	0.02	0.5	11%
	Correctness x Certainty	0.790	2.3	0.832	2.8	0.087	-1.71	0.22	0.5	21%
	Certainty only	0.923	7.0	0.923	7.8	0.984	-0.02	0.00	0.8	12%

TEST-2 Based on NATION (2009)
TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.829	5.5	0.829	6.5	0.003	-2.96	0.38	1.0	18%
	Correctness x Certainty	0.835	5.2	0.854	6.0	0.002	-3.05	0.39	0.8	16%
	Certainty only	0.851	8.2	0.877	9.0	0.043	-2.03	0.26	0.8	10%
6000-10000wd 25 items	Correctness only	0.766	1.0	0.744	1.0	0.008	-2.64	0.34	0.0	0%
	Correctness x Certainty	0.804	0.5	0.771	0.8	0.004	-2.87	0.37	0.3	67%
	Certainty only	0.853	2.0	0.860	2.0	0.116	-1.57	0.20	0.0	0%

TEST-3 Based on NATION (2009)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.809	8.0	0.809	9.0	0.635	-0.48	0.06	1.0	13%
	Correctness x Certainty	0.867	6.3	0.861	6.7	0.121	-1.55	0.20	0.3	5%
	Certainty only	0.916	11.0	0.910	14.3	0.036	-2.09	0.27	3.3	30%
6000-10000wd 25 items	Correctness only	0.731	3.0	0.578	3.0	0.726	-0.35	0.05	0.0	0%
	Correctness x Certainty	0.779	2.0	0.697	2.0	0.943	-0.07	0.01	0.0	0%
	Certainty only	0.940	7.0	0.917	7.0	0.339	-0.96	0.13	0.0	0%

TEST-4 Based on SCHMITT ET AL (2001)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.901	7.0	0.907	8.5	0.006	-2.76	0.35	1.5	21%
	Correctness x Certainty	0.938	3.7	0.941	6.2	0.000	-3.51	0.44	2.5	68%
	Certainty only	0.952	11.8	0.949	16.5	0.002	-3.05	0.38	4.7	39%
3000wd 30 items	Correctness only	0.881	4.5	0.885	6.0	0.000	-3.51	0.44	1.5	33%
	Correctness x Certainty	0.915	3.0	0.934	3.5	0.001	-3.29	0.41	0.5	17%
	Certainty only	0.961	10.0	0.953	13.5	0.102	-1.64	0.20	3.5	35%
Academic wd 30 items	Correctness only	0.897	5.0	0.907	7.0	0.002	-3.08	0.38	2.0	40%
	Correctness x Certainty	0.930	3.3	0.939	4.3	0.004	-2.85	0.36	1.0	30%
	Certainty only	0.946	11.2	0.960	13.3	0.189	-1.31	0.16	2.2	19%
5000wd 30 items	Correctness only	0.881	3.0	0.906	4.0	0.017	-2.40	0.30	1.0	33%
	Correctness x Certainty	0.930	1.5	0.945	2.0	0.020	-2.33	0.29	0.5	33%
	Certainty only	0.941	6.7	0.951	9.5	0.150	-1.44	0.18	2.8	42%
10000wd 30 items	Correctness only	0.736	1.0	0.836	2.0	0.117	-1.57	0.20	1.0	100%
	Correctness x Certainty	0.822	0.7	0.881	0.8	0.067	-1.83	0.23	0.2	25%
	Certainty only	0.956	4.8	0.967	6.0	0.098	-1.66	0.21	1.2	24%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill; Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill; Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{(N_{\text{pretest}} + N_{\text{posttest}})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = $((\text{posttest score} - \text{pretest score}) / \text{pretest score}) \times 100$. Note if pretest score = 0.0, then Percentage Change cannot be calculated.

Note: Less participants than the total control cohort of 36 participants are recorded, because a few of them were absent for some of the posttests, with these absentees being different individuals in different tests.

4.2.1 Internal Consistency

Cronbach's alpha coefficients were used to gauge inter-item correlation, and hence the internal consistency of the tests. The results of the Main Study (see Figure-7 p.139 & Appendix-18) demonstrated internal consistency between the items of the tests, with $\alpha > 0.7$ (Pallant, 2010:97) and these are highlighted in green. In many cases, the coefficients were considerably higher than 0.7, showing good internal consistency (*ibid.*, p.101). For example with the Experimental cohort, $\alpha > 0.8$ for Certainty Alone in the 1000-5000-word levels of Test-1 (Spelling) & Test-3 (Meaning), both of these tests being based on the Vocabulary Size Test (Nation, 2009) as detailed in Section-3.4.4 p.79. This was also the case for many results of the three most frequent word levels of Test-4 (Meaning) that was based on the Vocabulary Levels Test (Schmitt et al., 2001) as detailed in Section-3.4.5 p.86. Also, $\alpha > 0.8$ for the Control cohort in almost all cases of the 4 tests, showing even better internal consistency.

Cronbach's alpha coefficient is sensitive to the number of items tested (Pallant, 2010:97), which could explain why the sections in Test-4 ($n=30$) had higher values than the sections in the other tests ($n=25$).

In a few instances, poor internal consistency of the test items of the Experimental cohort results was found with $\alpha < 0.7$ (Pallant, 2010:97). These are highlighted in red in Figure-7 & Appendix-18, and it can be seen that this was especially the case for the less frequent 6000-10000-word levels of Test-2 (Open Translation) & Test-3 (Meaning), and the 10000-word level of Test-4 (Meaning). However, it should be noted that these were typically less than the 0.7 value by only a small margin, for example $\alpha = 0.662$ for the posttest of Test-3 for Correctness Factored by Level of Certainty in the 6000-10000-word levels. This possibly showed that even after the ER programme, the students were still finding these less frequent words difficult, which resulted in many blind guesses and omissions.

4.2.2 Results for the Different Aspects of Word Knowledge

The results given above in Figure-7, Figure-8 & Appendix-18 possibly suggest that with the tests that were based on the Vocabulary Size Test (Nation, 2009, as detailed in Section-3.4.4 p.79), the scores for Test-1 (Spelling) were higher than Test-3 (Meaning), which in turn were higher than Test-2 (Open Translation). For example, the Median Experimental cohort pretest scores for Correctness Alone for the 1000-5000-word levels were 10.0/25 for Test-1 which was more than 9.0/25 for Test-3 which was more than 7.0/25 for Test-2, while the corresponding Control cohort scores were 10.0/25, 8.0/25 and 5.5/25 respectively.

The Median pretest scores for Correctness Alone in Test-4 (Meaning) that was based on the Vocabulary Levels Test (Schmitt et al., 2001, as detailed in Section-3.4.5 p.86) for the Experimental cohort were 7.0/30 for the 2000-word level, 5.0/30 for the 3000-word level, and 3.0/30 for the 5000-word level, while the corresponding results for the Control cohort were 7.0/30, 4.5/30 and 3.0/30 respectively. This may suggest that the results for Test-4 can be approximately placed between Test-3 and Test-2 for those comparable word frequency levels.

Hence it could be cautiously implied that the results for Test-1 (Spelling) were higher than Test-3 (Meaning) which were more than Test-4 (Meaning) which were in turn more than Test-2 (Open Translation). This possibly showed that the participants had more receptive knowledge of spelling and meaning than productive knowledge of meaning at the pretest stage that was before the ER programme.

4.2.3 Results for the Different Word Frequency Levels

The above results in Figure-7, Figure-8 & Appendix-18 between the different word frequency levels suggest that for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning), the scores for the 1000-5000-word frequency levels were much higher than those for the 6000-10000-word levels. For example, the Median posttest scores for Test-2 (Open Translation) for Certainty Alone for the Experimental cohort were 10.7/25 for

the 1000-5000-word levels which was four times more than 2.7/25 for the 6000-10000-word levels. The Control cohort showed similar results with 9.0/25 for the 1000-5000-word levels and 2.0/25 for the 6000-10000-word levels.

For Test-4 (Meaning) the scores in the 2000-word, 3000-word and Academic-word levels seemed to be higher than the other levels. For example, the Median pretest scores for Test-4 for Correctness Alone for the Experimental cohort were 7.0/30 for the 2000-word level which was slightly more than 6.0/30 for the Academic-word level which was more than 5.0/30 for the 3000-word level which was in turn higher than 3.0/30 for the 5000-word level. The corresponding results for the Control cohort were 7.0/30, 5.0/30, 4.5/30 & 3.0/30 respectively. This was possibly expected, given that it was more likely that the participants would have better knowledge of words that appeared more frequently in both general English and academic English contexts. Once again, it was observed that the scores were modest, even in the more frequent word levels.

Hence overall it may seem that for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning), the results for the 1000-5000-word levels were much more than those of the 6000-10000-word levels. As for Test-4 (Meaning), the scores for the 2000-word level were more than the Academic-word level which was higher than the 3000-word level which were in turn more than the 5000-word level which were more than the 10000-word level. This could indicate that at the pretest stage, the participants had better knowledge of high frequency general and academic words than of the lower frequency 5000-word & 10000-word levels, which is usually required for reading unsimplified texts (Hirsh & Nation, 1992:689).

4.2.4 Results involving Levels of Certainty

The results in Figure-7 above, Figure-8 above & Appendix-18 involving different levels of certainty showed that Correctness Alone was always higher than Correctness Factored by Level of Certainty (as should have been the case), while the scores seemed to be highest for Certainty Alone. For example, the Median pretest scores for Test-2 (Translation) for the 1000-5000-word levels for the Experimental cohort were 8.7/25 for

Certainty Alone which was more than 7.0/25 for Correctness Alone which was greater than 6.0/25 for Correctness Factored by Level of Certainty, while the corresponding results for the Control cohort were 8.2/25, 5.5/25 & 5.2/25 respectively.

The small difference between Correctness Alone and Correctness Factored by Level of Certainty may indicate that the students were fairly sure about most of their correct answers. The higher Certainty Alone scores could indicate that the students were fairly certain about a significant number of words, but were still getting them wrong.

4.2.5 Distribution of Data

The descriptive statistics are presented in Appendix-18 and show that the standard deviations were sometimes large in comparison to the means, indicating a large spread of results, and even larger standard deviations were seen in the less frequent word levels. For example, the Experimental cohort's pretest scores for Test-4 (Meaning) for Correctness Alone were Mean=9.0/30 & Standard Deviation=5.7 at the 2000-word level, compared to Mean=3.4/30 & Standard Deviation=2.9 at the 5000-word level.

The descriptive statistics also showed that only few test score distributions were symmetrical. For example, the Experimental cohort's pretest scores for Test-2 (Open Translation) for Correctness Alone at the 1000-5000-word levels were Mean=6.8/25, Lower Quartile=4.0/25, Median=7.0/25 & Upper Quartile=9.0/25. However, the vast majority of data was significantly skewed. For example, the Experimental cohort's pretest scores for Test-4 (Meaning) for Correctness Alone at the 2000-word level were Mean=9.0/30, Lower Quartile=5.0/30, Median=7.0/30 & Upper Quartile=12.5/30. This skew was more pronounced in the less frequent word levels. For example, the Experimental cohort's pretest scores for Test-4 (Meaning) for Correctness Alone at the 10000-word level were Mean=1.8/30, Lower Quartile=0.0/30, Median=1.0/30 & Upper Quartile=3.0/30.

Overall, the descriptive statistics showed that the vast bulk of the data was not normally distributed, and so the non-parametric Wilcoxon Signed-Rank Test for repeated

measures was used to compare the pretests and posttests (Pallant, 2010:230) for statistically significant vocabulary gain.

4.2.6 Vocabulary Gain between Pretests and Posttests

As detailed above, the descriptive statistics showed that the data was not normally distributed, and so the hypothesis testing used to compare the pretests and posttests for statistically significant gain, was the non-parametric Wilcoxon Signed-Rank Test for repeated measures (Pallant, 2010:230).

These results above in Figure-6 p.138, Figure-7 p.139, Figure-8 p.140 & Appendix-18 showed that although there was an increase in the mean score in every case except one of the 33 pretest-posttest pairs, this increase was statistically significant ($p < 0.05$) in the more frequent 1000-5000-word levels of Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) that were all based on the Vocabulary Size Test of Nation (2009) detailed in Section-3.4.4 p.79, as well as in the 2000-word, 3000-word and Academic-word levels of Test-4 (Meaning) that was based on the Vocabulary Levels Test of Schmitt et al. (2001) detailed in Section-3.4.5 p.86. In addition, the significant gains with Test-4 (Meaning) were generally the most, with some of these being substantial in percentage terms i.e. $(\text{final score} - \text{initial score}) / \text{initial score} \times 100$, but very modest in terms of absolute numbers of words. The results involving the different levels of certainty (i.e. Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone) showed that none of these three showed consistently better statistically significant gains than the other two.

Most importantly, in reference to RQ1 concerning the impact of the ER programme on vocabulary acquisition (see Section-2.8 p.36), there was very little difference in statistically significant vocabulary gains between the Experimental cohort that followed the ER programme, and the Control cohorts that did not. In many cases, the gains of the Experimental cohort were more than those for the Control cohort, but only by a small margin, and in a few cases the Control cohort's gains were slightly more. Given these modest vocabulary gains and the very slight differences between them for the

Experimental and Control cohorts, it was deemed unnecessary to perform another analysis to see if these small differences in the change from pretest to posttest median scores between the cohorts were statistical significant.

Below are some examples of the statistically significant vocabulary pretest-posttest gains revealed by the Wilcoxon Signed-Rank Test for repeated measures in Figure-7 & Figure-8 above, after the Experimental Cohort received the ER programme for 13.1 weeks (see Section-3.2.5 p.66):

- Test-1 (Spelling) Certainty Alone, 1000-5000-word levels: the score for the Experimental Cohort increased significantly, $z=-4.54$, $p=0.000$ to 3 decimal places, with a medium effect size (Pallant, 2010:232) ($r=0.46$), from pretest Median=12.3/25 to posttest Median=14.3/25; while the score for the Control Cohort also increased significantly, $z=-1.97$, $p=0.049$, with a small effect size ($r=0.25$), but from pretest Median=13.0/25 to posttest Median=13.0/25 indicating a negligible gain of less than 0.05 words out of 25.
- Test-2 (Open Translation) Correctness Alone, 1000-5000-word levels: the score for the Experimental Cohort increased significantly, $z=-5.00$, $p=0.000$ to 3 decimal places, with a large effect size ($r=0.50$), from pretest Median=7.0/25 to posttest Median=9.0/25; while the score for the Control Cohort also increased significantly, $z=-2.96$, $p=0.003$, with a medium effect size ($r=0.38$), from pretest Median=5.5/25 to posttest Median=6.5/25, showing a smaller gain than the Experimental Cohort in terms of the absolute number of words.
- Test-3 (Meaning) Certainty Alone, 1000-5000-word levels: the score for the Experimental Cohort increased significantly, $z=-3.41$, $p=0.001$, with a medium effect size ($r=0.34$), from pretest Median=13.0/25 to posttest Median=14.3/25; while score for the Control Cohort also increased significantly, $z=-2.09$, $p=0.036$, with a small effect size ($r=0.27$), from pretest Median=11.0/25 to posttest Median=14.3/25, actually showing a larger gain than the Experimental Cohort in terms of the absolute number of words.
- Test-4 (Meaning) Correctness Alone, 2000-word level: the score for the Experimental Cohort increased significantly, $z=-4.39$, $p=0.000$ to 3 decimal places, with a medium

effect size ($r=0.44$), from pretest Median=7.0/30 to posttest Median=11.0/30; while score for the Control Cohort also increased significantly, $z=-2.76$, $p=0.006$, with a medium effect size ($r=0.35$), from pretest Median=7.0/30 to posttest Median=8.5/30, showing a smaller gain than the Experimental Cohort in terms of the absolute number of words.

- Test-4 (Meaning) Correctness Alone, 3000-word level: the score for the Experimental Cohort increased significantly, $z=-2.80$, $p=0.005$, with a small effect size ($r=0.28$), from pretest Median=5.0/30 to posttest Median=7.0/30; while the score for the Control Cohort also increased significantly, $z=-3.51$, $p=0.000$ to 3 decimal places, with a medium effect size ($r=0.44$), from pretest Median=4.5/30 to posttest Median=6.0/30, showing a slightly smaller gain than the Experimental Cohort in terms of the absolute number of words.

- Test-4 (Meaning) Correctness Alone, Academic-word level: the score for the Experimental Cohort increased significantly, $z=-2.18$, $p=0.029$, with a small effect size (Pallant, 2010:232) ($r=0.22$), from pretest Median=6.0/30 to posttest Median=7.0/30; while the score for the Control Cohort increased significantly, $z=-3.08$, $p=0.002$, with a medium effect size ($r=0.38$), from pretest Median=5.0/30 to posttest Median=7.0/30, actually showing a larger gain than the Experimental Cohort in terms of the absolute number of words.

4.2.7 Summary

Overall, the results above showed there were few statistically significant vocabulary gains after the ER programme of 13.1 weeks (see Section-3.2.5 p.66), and these were mainly in the more frequent word levels. It could also be cautiously suggested that the gains for Test-4 (Meaning) were higher than for Test-2 (Open Translation) which were in turn more than for Test-1 (Spelling), which were in turn greater than for Test-3 (Meaning). In addition, the results involving the different levels of certainty (i.e. Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone) showed that none of these three showed consistently better statistically significant gains than the other two.

Most importantly, in reference to RQ1 concerning the impact of the ER programme on vocabulary acquisition (see Section-2.8 p.36) after the ER programme of 13.1 weeks (see Section-3.2.5 p.66), the Experimental cohort showed greater statistically significant vocabulary gains than the Control in most of these cases, but the difference was not large in terms of absolute numbers of words, and in a few cases the Control actually showed slightly more vocabulary gain. A detailed discussion of these results is presented in Chapter-5 (Discussion).

4.3 END-OF-SEMESTER EXAM RESULTS

As detailed in the Methodology Chapter-3 (Section-3.1.2 Table-5 p.48), all participants of the Main Study (n=85) completed the institution's English Language End-of-Semester Exam during the examination weeks after the end of the course. Section-3.5 p.101 described the test, its administration and how it was used to see if the time taken up by the vocabulary tests and the ER programme had adversely affected the Main Pilot group's final examination results, and it reported that their results were actually significantly better than those of their whole batch, alleviating the fear in their case. In this section, the English Language End-of-Semester Exam results for the Main Study experimental and control cohorts are compared with those of their batch to check for the same, following the same methodology described in Section-3.5 p.101 for the Main Pilot group.

Excel spreadsheets and IBM's SPSS (v.17 and higher) were used to process the data to yield results for both the descriptive and inferential statistics (see Appendix-19). The descriptive statistics (Pallant, 2010:59-63) showed that the data was not normally distributed, and so the non-parametric Mann-Whitney U Test for 2 independent samples was used (*ibid.*, p.227), where $p < 0.05$ indicated significant difference between the groups (*ibid.*, pp.240-2), i.e. between the Experimental cohort, the Control cohort, and the whole of their batch.

4.3.1 Results

The above-mentioned Mann-Whitney U tests (see Appendix-19) showed that the End-of-Semester Exam results were significantly higher for the Experimental cohort (Median=83%, n=49) than for the total batch (Median=75%, n=691), $U=12490$, $z=-3.07$, $p=0.002$, $r=0.12$, while the results of the Control cohort (Median=78%, n=36) were not significantly different to the total batch (Median=75%, n=691), $U=12182$, $z=-0.21$, $p=0.835$. They also showed that the Experimental cohort's results (Median=83%, n=49) were not significantly different to that of the Control cohort (Median=78%, n=36), $U=698$, $z=-1.64$, $p=0.101$, noting too that beforehand at the start of the project, there

was also no significant difference between the Placement Test results of the Experimental cohort and the Control cohort (see Section-3.3 p.69).

All of this confirmed that, in terms of the institution's English Language End-of-Semester Exam results, the participants of the study were not disadvantaged.

4.4 READING DIARIES RESULTS

Chapter-3 (Methodology) Section-3.6 p.103 described that, in order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), reading diaries were used to record the reading of the participants during the ER programme. Section-3.6.1 p.104 detailed the items of the survey that were developed during the Initial Pilot & Main Pilot studies, with the final version for the Main Study being given in Appendix-12.

As briefly touched on in the Methodology chapter (see Section-3.6 p.103), the participants' responses needed a careful and consistent methodology of analysis to ensure the internal reliability of the research, such that other independent researchers could re-analysis the same data to yield similar results (Nunan, 1992:14-17). Hence some of the data needed to be operationalized (Burns, 1999:21-22), whereby they needed to be clearly defined and so easier to measure and hence easier to analyse. This involved categorizing responses and converting grouped options into discrete values, as is now given in detail below for the items of the Reading Diary, with reference to Appendix-12:

Item-1 (Week) required no special manipulation.

Item-2 (Language of Reading Material) required no special manipulation, being either English or Arabic.

Item-3 (Type of Reading Material) had 18 options (see Appendix-12 & Appendix-20), with Option-18-‘Other’ referring to instruction leaflets, user guides, manuals, catalogues, tourist guides, tickets, menus, posters, and other materials that did not fit into the other categories, like books with many pictures but with unsimplified language, and like books aimed for young people on wildlife, science and sport.

There were a number of specific points when dealing with these options. For Option-15-‘English Video + Arabic Subtitles’, Item-2 (Language of Reading Material) was

recorded as Arabic; for Option-4-‘Book with Translation’, Item-2 (Language of Reading Material) was recorded as English unless the student specified otherwise, because the students reported that they almost always used the Arabic translation just as a quick and easy-to-use dictionary; and for Option-12-‘Internet Text’, students often recorded Item-2 (Language of Reading Material) as both English and Arabic, and so in this case the duration of their reading was split accordingly (see below for Item-5).

In order to simplify these many types, they were also grouped into 5 broader categories or overall options: ‘College Material’ (Options-1,2,3), ‘Simplified Book Material’ (Options-4,5,6,7), ‘Unsimplified Book Material’ (Options-8,9,10,11,18), ‘Screen Text with No Video’ (Options-12,13) & ‘Video with Subtitles’ (Options-14,15,16,17).

Item-4 (Topic of Reading Material) was developed during the data-entry process, and great care was taken to note actual examples of each category, to maintain consistency throughout. Eventually there were 16 categories, with detailed examples being provided in Appendix-20.

Item-5 (Duration of Reading) had grouped intervals, and so the mid-values were used except for the extreme groups, which were based on what the participants described, as found out through the Supplementary Questioning (see Section-3.8 p.117 & Section-4.6 p.162). Hence (see Appendix-12) ‘Under 30min’ was recorded as 0.25-hours/week, ‘30min-1hour’ 0.75-hours/week, ‘1-2hours’ 1.5-hours/week, ‘2-3hours’ 2.5-hours/week, ‘3-4hours’ 3.5-hours/week, ‘4-5hours’ 4.5-hours/week & ‘Over 5hours’ 10-hours/week. When participants recorded reading in English & Arabic (e.g. when browsing the internet), Supplementary Questioning (see Section-3.8 & Section-4.6) was used to determine the approximate split. This item (duration of time spent reading) was used in preference to Yamashita’s (2004:5) choice of using the average number of pages read, deeming the later to be extremely difficult for the participants to gauge with any kind of accuracy or consistency.

Item-6 (Frequency of Reading) had discrete options except for ‘more than 7’ times a week which, following discussions with the participants (see Supplementary Questioning Section-3.8 p.117 & Section-4.6 p.162), was recorded as 10-times/week.

Item-7 (Ease of Reading) used a Likert scale of discrete options from 1 (very hard) to 5 (very easy).

Item-8 (Dictionary Use) had discrete options, but of differing units, so the entries were converted to have the same units, and values for the extreme options were based on what the participants described, as found out through the Supplementary Questioning (see Section-3.8 & Section-4.6). Hence (see Appendix-12) ‘>5wds/pg’ was recorded as 10-words/page, ‘2-5wds/pg’ 3.5-words/page, ‘1wd/pg’ 1-words/page, ‘1wd/2 pgs’ 0.5-words/page, ‘1wd/ 5pgs’ 0.2-words/page & ‘<1wd/5pg’ 0-words/page. Through the Supplementary Questioning (see Section-3.8 & Section-4.6) participants noted that their entry for this item was actually based on the approximate number of words they did not know, instead of actual dictionary use, because in many situations they did not have easy access to a dictionary, e.g. when watching videos with subtitles.

Item-9 (Enjoyment) & Item-10 (Interest) used Likert scales of discrete options from 1 (very ‘unenjoyable’) to 5 (very enjoyable) and 1 (very boring) to 5 (very interesting) respectively. As mentioned in the Methodology (Section-3.6.1 p.104), the subtle difference between the two was noted during the Initial Pilot, and then was given as an example during the Main Pilot and Main Study. This example was that of a news article that appeared at that time describing the gruesome murder of a young child, which was the kind of story that was not commonly found in the Saudi newspapers, and hence this article (which was very much the topic of conversation among the students at that time) was an example of reading that was not at all enjoyable due to the crime committed, but was nonetheless still interesting because it was so shocking.

Item-11 (How Essential the Reading was) used a Likert scale of discrete options from 1 (unessential) to 5 (really essential). As a bench mark, the participants were instructed to enter a score of ‘5’ for any reading they were forced to do, for example reading during

the in-class Sustained Silent Reading sessions (see Section-3.2.3 p.61), reading required to complete important assignments, and reading required to revise for tests.

4.4.1 Results

The results yielded from the Reading Diaries are given in Appendix-20, and salient points are summarized below, noting that all quantities for the items were weighted in terms of Item-5 (Duration of Reading).

Overall, for the Experimental cohort (n=49) of the Main Study, during the ER programme (13.1 weeks – see Section-3.2.5 p.66), a total of 5676 hours of English reading (8.84 hours/participant/week) and 1580 hours of Arabic reading (2.46 hours/participant/week) was recorded. These absolute amounts could be described as modest given the students were studying a university course, and they suggested that the students read much more English than their own L1.

As described above, Item-3 (Type of Reading Material) was analysed as 18 detailed options and as 5 overall options. Hence for the detailed options, Appendix-20 shows that by far the biggest single detailed type of English Reading was College Textbook (1917 hours i.e. 2.99 hours/participant/week), followed by Internet Text (1189 hours i.e. 1.85 hours/participant/week). The largest amounts of Arabic reading were of Internet Text (671 hours i.e. 1.05 hours/participant/week) and English Movies with Arabic Subtitles (632 hours i.e. 0.98 hours/participant/week). In terms of the overall options, Appendix-20 shows that the largest amount of English reading was of College Material (2433 hours i.e. 3.79 hours/participant/week) and Screen Text with No Video (1548 hours i.e. 2.41 hours/participant/week), and for Arabic it was for Screen Text with No Video (684 hours i.e. 1.07 hours/participant/week) and Video with Subtitles (635 hours i.e. 0.99 hours/participant/week).

The results for Item-4 (Topic of Reading Material) showed that by far most of the English reading was College Related (2639 hours i.e. 4.11 hours/participant/week), while Arabic reading mainly related to Texts/Emails (364 hours i.e. 0.57

hours/participant/week), Action/Adventure/Thrillers (289 hours i.e. 0.45 hours/participant/week), News (210 hours i.e. 0.33 hours/participant/week) and Dramas/Soap Operas (189 hours i.e. 0.29 hours/participant/week).

The results for Item-6 (Frequency of Reading) showed that most of the English reading was done 10, 5 and 4 times/week, i.e. approximately every week day or more, perhaps relating to how often the participants would need to read their college-related material. Arabic reading had more evenly distributed frequencies.

The results for Item-7 (Ease of Reading) seemed to relate well with those of Item-8 (Dictionary Use), given that about two-thirds of the English reading was rated 4/5-5/5 (with 5/5 being 'very easy'), and about two-thirds of the dictionary use was at a rate of just 0-1 words/page.

The results for Item-9 (Enjoyment) and Item-10 (Interest) (see above and Section-3.6.1 p.104 for the subtle difference between the two), also seemed to correspond, given that for Item-9 more than two-thirds of the English reading was rated 4/5-5/5 (with 5/5 being 'very enjoyable'), and for Item-10 almost three-quarters was rated 4/5-5/5 (with 5/5 being 'very interesting').

The results for Item-11 (How Essential the Reading was) showed that most (about two-thirds) of the English reading was rated 4/5-5/5 (with 5/5 being 'really essential').

It should be noted, as first mentioned in Section-3.1.3 p.50, that statistical correlation analyses using multiple regression and structural equation modelling were considered to quantify the relationship between these characteristics of the students' reading and their vocabulary acquisition during the ER course (given in the Results Section-4.2.6 p.145). However the researcher was not explicitly seeking to gain a simple generalisation that could be offered by such correlation analyses, and instead the aim was to provide an in-depth picture of the participants in this specific context. Hence it was decided to give these results with all their details, as done so above, to preserve the very human and

social nature of this particular part of the project, and to provide a rich description that could potentially contribute more to Chapter-5 (Discussion).

4.4.2 Summary

In summary, these results provided a profile of the participants' reading during the ER programme, and so could help to shed light on why the subsequent vocabulary gains were modest (see Section-4.2.7 p.147). Firstly, the participants read little even in Arabic (2.46 hours/participant/week – see above), suggesting that reading itself was a skill they were not used to, and further suggesting that reading in the foreign language of English could be extremely challenging. Also, Item-3 (Type of Reading Material) and Item-4 (Topic of Reading Material) showed that the overwhelming majority of the English reading was college-related, and this was possible reflected in Item-11 (How Essential the Reading was) that indicated that most of the English reading was ranked as 'really essential', even though Item-7 (Ease of Reading), Item-8 (Dictionary Use), Item-9 (Enjoyment) & Item-10 (Interest) seemed to indicate that a significant amount of the English reading was still 'very easy', 'very enjoyable' and 'very interesting'. A detailed discussion of these results is presented in Chapter-5 (Discussion).

4.5 SURVEY OF EXPOSURE TO ENGLISH LEARNING RESULTS

Chapter-3 (Methodology) Section-3.7 p.111 described that, in order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), a survey was administered to gauge the students' exposure to sources of English beyond the ER programme and the English Language course, in an attempt to address the limitations of some previous studies, such as Hafiz & Tudor (1989) that did not account for such exposure. Section-3.7.1 p.112 detailed the items of the survey that were developed during the Initial Pilot & Main Pilot studies, with the final version for the Main Study being given in Appendix-13.

As briefly touched on in the Methodology chapter (see Section-3.7 p.111), the participants' responses needed a careful and consistent methodology of analysis to ensure the internal reliability of the research, such that other independent researchers could re-analysis the same data to yield similar results (Nunan, 1992:14-17). Hence some of the data needed to be operationalized (Burns, 1999:21-22), whereby they needed to be clearly defined so that could be easier to measure and hence analyse. This involved categorizing responses and converting grouped options into discrete values, as is now given in detail below for the items of the Survey of Exposure to English Learning, with reference to Appendix-13:

Item-1 (Age) required no special manipulation.

Item-2 (Formal Study of English Outside the Current College) had 3 categories with further sub-categories. The first category was Level, which was sub-divided into School, College, University, Language Institute, Work/Training & Home Self-study. The next was Type, which was sub-divided into State Funded Education, Private Education & Self-study. Finally, there was the Place of the formal study, which, for these participants, was only in Saudi Arabia or abroad in the Far East. In terms of duration of study that was taught in English, the main unit was hours. Also, 1 month was taken as 4 weeks, and 1 year was taken as 30 weeks, because the system throughout the country is a total of 30 teaching weeks during the academic year (2 semesters of 18

weeks, each with 1 revision week and 2 exam weeks). The durations of each entry were multiplied by the percentage of time the lesson was taught in English, as recorded by the participants in the 6th column of the item (see Appendix-13).

Item-3 (External English Exams) yielded 3 categories: Company Entrance Test, School Leaving National Exam and College/University Entrance Exam. All the scores were converted into percentages for easy comparison.

Item-4 (Travel Abroad) had the durations (converted into days) of each entry multiplied by the percentage of time English was used, as recorded by the participants in the 5th column of the item (see Appendix-13). The entries yielded the following categories of Place: BANA (Britain, Australia and North America), Europe excluding the UK, Turkey, Arabic Countries & the Far East; and they yielded the following categories of Purpose: Tourism, Training & Medical Care.

Item-5 (Frequency of Watching English) and Item-7 (Frequency of Listening to English) had options that were simplified according to what the participants described, as found out during the Supplementary Questioning (see Section-3.8 p.117 & Section-4.6 p.162). Hence (see Appendix-13) ‘More than once a day’ was recorded as 60-times/month, ‘Once a day’ 30-times/month, ‘More than once a week’ 10-times/month, ‘Once a week’ 5-times/week, ‘More than once every 2 weeks’ 3-times/week, ‘Once every 2 weeks’ 2-times/week, ‘More than once a month’ 1.5-times/week, ‘Once a Month’ 1-time/month & ‘Less than once a month’ was recorded as 0-times/month.

Item-6 (Time Watching English) and Item-8 (Time Listening to English) had grouped intervals, and so the mid-values were used except for the extreme groups, which were based on what the participants described, as found out during the Supplementary Questioning (see Section-3.8 & Section-4.6). Hence (see Appendix-13) ‘Less than 30 min a week’ was recorded as 0.25-hours/week, ‘30min to 1 hour’ 0.75-hours/week, ‘1-2 hours’ 1.5-hours/week, ‘2-3 hours’ 2.5-hours/week, ‘3-5 hours’ 4-hours/week, ‘5-10 hours’ 7.5-hours/week, ‘10-20 hours’ 15-hours/week & ‘More than 20 hours a week’ was recorded as 30-hours/week.

Item-9 (Work Experience) had the durations (converted into hours) of each entry multiplied by the percentage of time English was used, as recorded by the participants in the 6th column of the item (see Appendix-13). The entries yielded the following categories: Civil Service, Teaching, Hospital, Fast Food Restaurant, Hotel, Own Business, ICT, Construction, & Bank.

4.5.1 Results

The results yielded from the Survey of Exposure to English Learning are given in Appendix-21, and salient points are summarized below.

The results for Item-1 (Age) showed the participants were predominantly in their early twenties (n=49, Median=22yrs), although 3 of them were in the mid-thirties, which affected some of the descriptive statistics (Mean=23yrs, Standard Deviation=3).

The results for Item-2 (Formal Study of English Outside the Current College) showed that 66 per cent of their learning of English in English was at the college they attended prior to the current university (n=49, Mean=595hours, Median=618hours), which was the first time most of them were taught by English-speaking teachers, and 95 per cent of their English was learned at the formal educational setting of school, college and university. In addition almost all (94 per cent) of their English learning was in state-funded institutions, and practically all (>99 per cent) was done within Saudi Arabia.

The results for Item-3 (External English Exams) showed that 13 of the 49 participants had taken external English exams, and some had taken more than one. The largest category was Company Entrance Test (54 per cent), followed by College/University Entrance Exam (33 per cent). The Mean scores were approximately 60-70 per cent, noting that 90 per cent is the A grade boundary in the Saudi Arabian educational system. Supplementary Questioning (see Section-3.8 p.117 & Section-4.6 p.162) of these students revealed that none undertook any special preparation for the exams.

The results for Item-4 (Travel Abroad) showed that 18 of the 49 participants had used English abroad. Few students had travelled to BANA (n=3, Mean=22days), while most had travelled to Arabic countries, but used little English during these trips (n=14, Mean=29days), perhaps because all of these happened to travel for Tourism only. The few who had travelled to the Far East used more English (n=4, Mean=59days), noting that most of these (n=3) happened to travel there for training in the English medium, and Supplementary Questioning (see Section-3.8 p.117 & Section-4.6 p.162) revealed these were the older students aged over 30.

The results for Item-5 to Item-8, pertaining to the frequency and amount of English watched and listened to, showed widely varying figures for the participants (n=49), as readily shown by the histograms, with large differences in means and medians, and large standard deviations (see Appendix-21). For example, Item-5 (Frequency of Watching English) had Mean=26 times/month, Median=10 & Standard Deviation=24, and Item-6 (Time Watching English) had Mean=7 hours/week, Median=4 & Standard Deviation=9. Supplementary Questioning (see Section-3.8 & Section-4.6) revealed that much of what they watched were American action movies, often with Arabic subtitles. Item-7 and Item-8 showed they listened to English approximately half as much as they watched videos.

The results for Item-9 (Work Experience) showed that 18 of the 49 participants had used English at work. A few of these had used English during their full time work in the Civil Service before coming to the current university, which represented the largest use of English (n=5, Mean=3680hours), with Supplementary Questioning (see Section-3.8 & Section-4.6) revealing that these were the older students. Others had used English in many other fields, including in their own business (n=2, Mean=1800hours), in construction companies (n=1, Mean=1560hours) and in fast-food restaurants (n=1, Mean=920hours).

It should be known, as first stated in Section-3.1.3 p.50, that statistical correlation analyses using multiple regression and structural equation modelling were considered to quantify the association between these features of the students' English learning given

above, and their vocabulary acquisition during the ER course that was given in the Results Section-4.2.6 p.145. However the researcher was not explicitly seeking to gain a simple generalisation that could be offered by such correlation analyses, and instead the goal was to offer a thorough description of the students in this particular situation. Therefore it was decided to give these results with all their details, as done so above, to maintain the very human and social nature of this specific section of the study, and to provide a detailed account that could potentially contribute more to Chapter-5 (Discussion).

4.5.2 Summary

Overall, these results provide a profile of the participants' exposure to English outside their current university, and so may help to shed light on why the vocabulary gains following the ER programme were so modest (see Section-4.2.7 p.147). Item-1 (Age) and Item-2 (Formal Study of English Outside the Current College) showed the participants were predominantly in their early twenties, coming straight from college, which was the first time most had been taught English in English with native English speaking teachers, and almost all of their academic learning was in Saudi Arabia, where English is not used much outside the formal education setting. Item-3 (External English Exams) showed that few had taken external exams, none had made any special preparation for them, and their scores were the equivalent to C-D grades in the Saudi Arabian educational system. Item-4 (Travel Abroad) indicated that most had travelled for tourism, and few had travelled outside the Arab world, resulting in little English being used during such trips. The few that had used a lot of English were the older participants aged over 30, and they had been sent for training in English. Item-5 to Item-8, pertaining to the frequency and amount of English watched and listened to, showed many participants watch English videos every day, but mainly American movies, often with Arabic subtitles. Finally Item-9 (Work Experience) showed that few had any work experience, and that it was the older participants who had used the most English in the civil service before coming to the university. A detailed discussion of these results is presented in the Chapter-5 (Discussion).

4.6 SUPPLEMENTARY QUESTIONING RESULTS

In order to address both RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), hundreds of supplementary questions were posed throughout the study to the participants, and their responses were used to gain a deeper understanding of what they recorded in the data collection instruments, as recommended by Hyland (2002:166) and many others. As detailed in Section-3.8 p.117, the questions were posed in a systematic manner when checking entries in the data collection instruments, and at other times they were posed to immediately clarify strange or interesting entries. These clarifications were immediately written down on that particular student's data collection instrument and in the Research Journal (see Section-3.10 p.129). By recording the answers immediately, accuracy was ensured as much as possible (as pointed out by Benson, 2001:203), and by making the questions informal, the students were as relaxed as possible when answering. The resulting information was analysed with emphasis on 'meanings, experiences . . . [and] descriptions' (Swann, 1994:29), taking the view that a qualitative subjective approach would be more appropriate, given the spontaneous and personal nature of the raw data.

4.6.1 Results

Hence, as already detailed in many previous sections, the 'results' of this data collection instrument, i.e. the answers to the Supplementary Questioning, shaped the development of the whole study. During the Initial Pilot, the answers to the Supplementary Questioning posed to the focus groups moulded the initial versions of the data collection instruments and the ER programme (Section-3.1.1 p.42). One example of this was concerning the in-class Sustained Silent Reading sessions (Section-3.2.3 p.61) which as a result of the answers to the Supplementary Questioning, were modified for the Main Pilot and Main Study in terms of duration and timing within the lessons. Another example was during the development of the vocabulary tests, when the answers to the Supplementary Questioning led to important amendments to the Likert Scale used to gauge certainty with each answer in each test (Section-3.4.3), to the items in Test-1

(Spelling) & Test-3 (Meaning) (Section-3.4.4.1 p.82, Section-3.4.4.3 p.85 & Appendix-5), to the layout of Test-4 (Meaning) (Section-3.4.5 p.86) and to the administration of the tests (Section-3.4.6 p.89).

The answers to the Supplementary Questioning continued to have an important role, even during the analysis of the data from the other data collection instruments. For example, the answers to the Supplementary Questioning determined what discrete values were used when analysing some grouped interval responses in the Reading Diaries (Section-4.4 p.151) and the Survey of Exposure to English Learning (Section-4.5 p.156), and for the latter they were also used to gain more in-depth information about the responses (Section-4.5), like the kind of special preparation taken before taking External English Exams, the nature of the participants who had travelled abroad for training in the English medium, and the kind of English videos participants watched.

The answers to the Supplementary Questioning (with the findings of the other data collection instruments) also helped to form the basis of the direct questions used during the In-Depth Interviews (Section-3.9.2 p.120 & Section-4.7 p.164), aimed at probing into the factors that could explain the participants' reading habits and vocabulary gains during the ER programme.

4.7 IN-DEPTH INTERVIEWS RESULTS

In order to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), in-depth interviews were conducted as detailed in Section-3.9 p.118. These were used to try to ascertain the possible factors that may explain the vocabulary acquisition revealed in the vocabulary tests (see Section-3.4 p.72 & Section-4.2 p.135). Hence the quantitative data (i.e. the vocabulary tests, reading diaries & surveys) was principally used to describe participants in terms of *what* they were like, while the qualitative data (i.e. the Supplementary Questioning & In-Depth Interviews) was largely used to seek *why* they were like that. It should be pointed out that even though the quantitative data was also exhaustive and detailed and could be used to infer some of these reasons, the qualitative data would divulge more of them directly from the words of the students themselves. As a result, this methodology follows up quantitative data with qualitative data, which Benson (2001:196) reports to be common in research because using a diverse selection of viewpoints, approaches and sources of data can greatly improve the authority of the subsequent results (Chappelle & Duff, 2003:165).

As briefly touched on in the Methodology chapter (see Section-3.9 p.118), a careful and consistent methodology was needed to analysis the interviews, in order to improve the internal reliability of the research, such that other independent researchers could re-analysis the same data to yield broadly similar results (Nunan, 1992:14-17). Qualitative data requires such appropriate and robust methods of analysis in order to enable researchers to explain how their claims emerged from their data (Richards, 2003:4), and this is now given in detail at the beginning of this section.

Hence to provide a precise methodology to analyse the data, a thematic analysis was adopted, being considered ‘best suited to elucidating the specific nature of a given group’s conceptualisation of the phenomenon under study’ (Joffe, 2012). Thematic analysis is a process used to identify, analyse, and report themes within data (Braun & Clarke, 2006), with Aronson (1995) stating that the themes normally focus on ‘patterns of living and/or behaviour’. Thematic analysis was considered appropriate for this

research because it is not viewed to be theoretically bound (Braun & Clarke, 2006), as it can be performed from a realist and/or constructionist viewpoint (Vaismoradi et al., 2013:399), and it can also allow for a mix of both deductive research question driven analysis and inductive data-driven analysis, as performed and described in detail by Fereday & Muir-Cochrane (2006), and as adopted here and described below in this section.

When performing the thematic analysis, the 6 phases described by Braun & Clarke (2006) were followed, namely:

1. Getting to know the data;
2. Generating initial codes;
3. Searching for themes;
4. Reviewing themes;
5. Defining and naming themes;
6. Producing the report.

As pointed out by Fereday & Muir-Cochrane (2006:83), this process was not linear, but iterative and reflexive, as detailed below.

Phase-1 involved getting to know the data, in order to be aware of its depth and breadth. Bradley et al. (2007:1761) described this as ‘Immersion in the data’ to impress on how thorough this should be to comprehend the data in its entirety. The data was the transcriptions of the recorded interviews of the participants that were typed up into Microsoft Word (see Section-3.9 p.118), and so as recommended by Braun & Clarke (2006), the transcription process was itself part of Phase-1, during which the data was ‘immersed’ into. Transcriptions of recordings are generally agreed to be more accurate than note-taking during the interview (Dörnyei, 2007:139), but this better accuracy can only be realized if the transcriptions themselves are detailed.

With this in mind, the transcriptions were verbatim, as shown in Appendix-22, which gives the full transcript of the interview with Interviewee-09. This shows the inclusion of repetitions e.g. *‘I read... one book is very very very big this book, but not complete.’* (Transcript-Interviewee-09-p.11), grammatical mistakes e.g. *‘I think you have err, you*

can give the student choose what he want... (Transcript-Interviewee-09-p.3), interjections e.g. *'Ah, yes...'* (Transcript-Interviewee-09-p.10), hesitations e.g. *'For me erm, for remember some vocabulary and so on...'* (Transcript-Interviewee-09-p.7) and non-verbal expressions e.g. *'Err it's good, but if you bring tea or coffee [laughing] that good.'* (Transcript-Interviewee-09-p.3), to preserve as much information as possible,

Hence this Phase-1 of getting to know the data started during the transcription process, and once completed, the transcripts were read again to gain a further familiarity with their contents.

Phase-2 involved generating Initial Codes. Fereday & Muir-Cochrane (2006:83) described coding as recognizing 'an important moment' and then assigning a description that encapsulated the qualitative depth of it. Hence in this analysis, the Initial Codes were simply short phrases that described the import of any particular transcribed text 'extract' that appeared 'relevant' or 'interesting', all of which are defined below.

To define the term 'extract', Aronson (1995) suggested that either direct quotes or paraphrasing common ideas could be used to determine patterns in the data. However, to be more precise and to keep a close link to the data, direct quotes were preferred as the basis of each 'extract', the length of which varied according to what was required to preserve the meaning and context when isolated into a quote, and if they were replies to a question, the question was also kept with the 'extract'. Hence most extracts were a sentence, for example *'To be frank, reading is a problem, if in English or Arabic.'* (Transcript-Interviewee-04-p.11), while others were simply one word replies with the question also added, like *'Yes.'* (Transcript-Interviewee-01-p.8), when asked if young people don't like reading books. However, some extracts were longer like

'No because, most of my colleagues they even have a problem in Arabic. You know what's the difference between ض and ظ! [Laughing]. They still have a problem with this, even in Arabic, because they didn't read you know. Just for school and that's it. I think it is very hard.'

(Transcript-Interviewee-05-p.9).

All of these example extracts were given the Initial Code ‘ER Culture is Lacking’, and during the description of Phase-4 below, more extracts are given to serve as detailed examples of how the full coding process was performed.

The term ‘relevant’ referred to anything that was clearly connected to the aim of the In-Depth Interviews, i.e. to address RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), and hence anything relating to possible factors that may explain the modest vocabulary acquisition revealed in the vocabulary tests (see Section-4.2 p.135). These ‘relevant’ extracts allowed for a more deductive approach to the analysis, linking the research question to the data. Some of these were latent or explicit (Braun & Clarke, 2006) e.g. *‘I don’t have time, so I can’t read. You know the last semester, 32 hours.’* (Transcript-Interviewee-09-p.7 & Appendix-22). This was given the Initial Code ‘Work Load at College’. However, others had a semantic or interpretative relevance (Braun & Clarke, 2006), like

‘This is the truth about it. This is most difficulties I face with my studies, the number of classes. Not about the teacher, because we face, we had classes with many teachers. Not about the teachers themselves, but about the number of classes.’

(Transcript-Interviewee-05-p.1).

This was also given the Initial Code ‘Work Load at College’, because although the interviewee was talking about difficulties with his overall studies, this was interpreted to also affect his ability to engage in ER and/or benefit from it.

The term ‘interesting’ simply referred to anything that was new, unexpected, or potentially relevant. Such an extract need not have been directly related to RQ2, but could have simply triggered a new thought in the mind of the researcher. For example, one extract was

‘First job, er, I want er, Tahoe, Chevrolet Tahoe.... Of course new! In cash, it depends on the normal or extras... The normal 100, maybe 170000 [i.e. about 28000 GBP]’

(Transcript-Interviewee-10-pp.10-11).

The researcher was surprised that the student wanted an expensive brand new sports utility vehicle, making it clear how important it was to get a well-paid job upon graduation. This was given the Initial Code ‘Jobs are Very Important’. Other extracts were noted for giving information that was probably new to those who are unfamiliar with this context. An example of this was

‘Yes, the wedding hall maybe 30000 [i.e. about 5000 GBP] is one night. For the dinner maybe also 30000, and 30000 for the wedding gift for the woman. And you need small, er, apartment ... Yes, you need spend like this and more. This is simple marriage.’

(Transcript-Interviewee-09-p.10 & Appendix-22).

This described the high cost of getting married (which the interviewee wanted to do two years after graduation). This was also given the Initial Code ‘Jobs are Very Important’. Hence this attention to ‘interesting’ points served to give a wider context for the students’ voices that could become more relevant later on. It also allowed the analysis to be more inductive in nature, linking the data up to potentially new ideas, agreeing with Joffe’s (2012) view that ‘there is little point in conducting qualitative work if one does not want to draw on the naturalistically occurring themes evident in the data itself.’

As can be seen from the examples above, attention was given to both the extracts that described the experiences of the participants, and to the extracts that described their society. This allowed for a more contextualized analysis, which accounted for both the role of the individuals and the role of the society in which they are placed, viewing both to be significant and interconnected. All of this shows the flexibility that thematic analysis can offer, which was one the primary reasons it was chosen for this project.

Hence in this Phase-2, as much as possible was coded, and as little as possible was discarded. In many instances, the same extract was given multiple Initial Codes, leading to an expansion to the list of ‘relevant’ and ‘interesting’ ‘extracts’. For example

‘Yes, have grades. This is the first thing to work. Give grades in the test or the vocabulary, they will work...’

(Transcript-Interviewee-10-p.12).

This was given the two Initial Codes: ‘Discipline, Scaffolding & Follow-Up are Needed to Learn’ and ‘Grades are Very Important’.

The many examples above illustrate that the coding was not always clear-cut, which was not surprising, given the natural nature of the data. They also demonstrate some of the detailed considerations that were taken to find the most appropriate Initial Codes, and further details of this process are given below when describing Phase-3 and Phase-4 in which detailed worked examples are given.

In order to mechanically process this phase, the Initial Codes were set up in Microsoft Excel (see Figure-9 below), with each associated extract of transcribed text being copied from the Word document and pasted next to it. Hence the spread sheet had the following 6 columns: ‘Code’, ‘Interviewee’, ‘Extract’, ‘Context’, ‘Notes’, ‘Source’. It was important, as touched on earlier, to retain a complete picture of each extract, hence necessitating the other columns. Sometimes it was necessary to note the question or general context associated with the extract, for example, ‘Yes’ (Transcript-Interviewee-01-p.8) was entered into Column-3 ‘Extract’, but next to it in Column-4 ‘Context’ ‘When asked if young people don’t like reading books’ was also added. Column-5 ‘Notes’ allowed for an extra comment to be added. For example an application called Moment was described to ‘...give you... how much time you spent on reading of Whatsapp, on Twitter, on er any website.’ (Interview-06-p.4), and so the comment ‘Excellent idea to record reading time’ was added next to it in Column-5. Finally, Column-6 gave the original file of the transcript, for easy cross-reference if needed. The spreadsheet layout with the two examples described above, are given in Figure-9 below.

Figure-9: Layout of the Excel Spreadsheet used in Phase-2 to assign Initial Codes to Extracts, showing the two example extracts detailed in the main text.

INITIAL CODE	INTERVIEWEE	EXTRACT	CONTEXT	NOTES	SOURCE
ER Culture is Lacking	Interviewee-01	Yes	When asked if young people don't like reading books		Interview xxxxxx Xxxx Xxxx XXXXXXXX XXXXXXXX XXXX.docx
English Learning Beyond Classroom	Interviewee-06	I don't like wasting a lot of time. Another thing the other day actually I was reading about a new application on iOS on iPhone, named Moment or something. That Moment application can access and analyse your action with the phone. That he can give you in the end of the day an information about how many er you know, how much time you spent on reading of Whatsapp, on Twitter, on er any website. It just figure that out from a behaviour. So you know I was surprised I spent 7 hours of reading!?! I didn't get a lot of information. It was a video website, it was Whatsapp, it was public information about Abha, it was Twitter.		Excellent idea to record reading time	Interview xxxxxx XXXXXXXX.docx

The use of dedicated qualitative analysis computer software was considered during this and the following phases. Software such as ATLAS.ti, NUD*IST, NVivo and others have been considered useful when processing large amounts of interview data, but it should be emphasised that these packages ‘cannot analyse textual data in the way that they can numerical data’ (Joffe, 2012). Hence a view was taken that searching for key words and counting codes is meaningless without accounting for the subtle nuances of meanings (Hewitt-Taylor, 2001:41), emphases, intonations, and other details that only the human eye, ear and mind can fully notice. Thus the decision was made to analyse the data by hand and with Excel, with this methodology being considered more thorough and sensitive, and allowing the researcher to get a greater first-hand understanding of the whole process.

Overall during Phase-2, over 920 extracts were used from the interviews, and they were assigned to the 40 Initial Codes that are listed in Appendix-23.

Phase-3 involved searching for Potential Themes. In this phase, the analysis was taken to a broader level, taking the Initial Codes from Phase-2, and sorting them into Potential

Themes. This was done by first rearranging the Initial Codes into Main Codes and Sub-Codes. The Main Codes would then be likely to form the basis of eventual themes.

‘Themes’ can be thought of as ‘patterns of explicit and implicit content’ (Joffe, 2012), or ‘general propositions that emerge... and provide recurrent and unifying ideas regarding the subject of inquiry’ (Bradley et al., 2007:1766). Braun & Clarke (2006) added that a theme ‘captures something important about the data in relation to the research question, and represents some level of *patterned* response or meaning within the data’, but also made it clear that there are no definitive rules for what that constitutes. The important point is that however a theme is defined, then that definition is applied consistently throughout the analysis. Vaismoradi et al. (2013:400) reported the view that thematic analysis should be used to search for and identify common threads across a whole set of interviews. However, given the expectation that such a stipulation of unanimity may only result in very few general themes, it was decided that in the thematic analysis of this data, a ‘theme’ was taken to mean a patterned response or meaning that was voiced by at least 9 of the 12 interviewees, i.e. by 75 per cent or more of the interviewees, thus hopefully allowing for more themes that would offer a richer description of the data. As suggested by Aronson (1995), the themes could be of many forms, such as topics of conversation, specific words, specific phrases, habits, meanings, emotions and feelings. With these criteria, and given that the interviewees were carefully chosen to be as representative as possible of the rest of the participants in the experimental cohort (see Section-3.9.1 p.119), the resulting themes could be considered a basis for future conclusions and recommendations.

In order to rearrange the Initial Codes into Main Codes and Sub-Codes, and hence work towards the Potential Themes, the spreadsheet formed during Phase-2 was kept, and a copy of it was made for the manipulation required in Phase-3, hence removing the chance of unwittingly deleting any data. The ‘Sort’ function in Excel was used to gather all identical codes together, and then the main process of cutting and pasting rows of Initial Codes and their associated extracts into Main Codes and Sub-Codes started. At this stage, no codes were discarded, but instead the aim was to just improve the organization. Also some extracts were given different or additional codes to those

assigned in Phase-2, but to a much lesser extent than during Phase-2. This was an example of how the different phases were not completely distinct from each other, but instead the process was recursive as advised by Braun & Clarke (2006) and 'iterative' as Bradley et al. (2007:1760) described many methods of qualitative analysis to be.

Appendix-24 shows the improved arrangement of Initial Codes into Main Codes and Sub-Codes during this Phase-3 that aided the emergence of Potential Themes. For example, the Main Code 'ER Culture is Lacking' had 5 Sub-Codes: 'Abstract', 'ER is Not Enjoyable', 'ER is Difficult', 'ER Lacking When Young', and 'ER of Books is Not Enjoyable'. For this particular Main Code, all of its above Sub-Codes were not Initial Codes during Phase-2, but were introduced at this stage to add detail to the Main Code. In this case, the Main Code 'ER Culture is Lacking' emerged as a Potential Theme, as detailed in Phase-4 (see below). The Sub-Code 'Abstract' was used as the first Sub-Code in many other Main Codes as well, for extracts that expressed the Main Code in a general or introductory manner.

The above demonstrated that the search for themes was not always clear, which was not unexpected, given the natural nature of the data. It also demonstrated some of the meticulous considerations that were taken to find the most appropriate themes, further details of which are given with the worked examples that are given below when describing Phase-4.

The 'CountIfs' and 'Sort' functions of Excel were used to list the Main Codes and Sub-Codes in order of occurrence (see Appendix-24), but not as a definitive way to determine which were more important than others. Instead this was just a convenient way to gauge an overall understanding of the new arrangement of codes, by considering, but not relying on, what Joffe, (2012) describes as their 'manifest content', i.e. how often they occurred, and instead agreeing with the view that the importance of a code does not always depend on quantifiable measures, but instead on whether it covers something important (directly or indirectly) to the research question (Vaismoradi et al., 2013:403). This phase yielded some frequently occurring Main Codes (see Appendix-24), like 'ER Culture is Lacking' (94 extracts) and 'Grades are Very Important' (77),

which were clear candidates for Potential Themes. Some Main Codes were less frequent, e.g. 'Jobs are Very Important' (42 extracts) and 'Family Ties & Responsibilities are Strong' (36). However, about 20 Initial Codes from Phase-2 did not seem to easily fit within the new arrangement of Phase-3. These were not ignored at this stage, especially given that some had a significant number of extracts, e.g. the code 'Pleasing or Impressing the Teacher' had 10 extracts. These were instead addressed in Phase-4.

Phase-4 entailed reviewing themes by revising the Main codes of Phase-3 that were Potential Themes. This was done by reading them again in detail, with their extracts. The key concern, as advised by Braun & Clarke (2006), was to check that there was a coherent pattern between themes and extracts. Then the Main Codes and Sub-Codes of Phase-3 were rearranged further to fit within the Potential Themes that were emerging. Once again, the recurrence of a theme was of some consideration, but not the only consideration, agreeing with Joffe (2012) who pointed out, 'the prevalence of a given theme does not tell the whole story'. Hence a further process in this Phase-4 was that careful consideration was given to how many interviewees voiced the Potential Theme, noting, as defined above when describing Phase-3, that a 'theme' should represent a patterned response or meaning that was voiced by at least by 75 per cent or more of the interviewees. Also, the beginning of a 'narrative' was mapped out, to describe the relationship between the themes, which is further detailed below in Phase-5 and Phase-6. All of this was again a recursive process, which required reference to what was found during the previous phases, but to a much lesser extent than before. Appendix-25 gives a full list of these Potential Themes, including how they were mapped from the Main Codes and Sub-Codes that were generated during Phase-3.

Like during the previous phases, some regrouping of the codes was done. Some Main Codes and Sub-Codes were moved, and some of the Initial Codes from Phase-2 that did not fit into a Main Code during Phase-3 were also moved. For example, one code 'Reading Only for Academic Studies', which during Phase-3 was a Sub-Code that did not seem to fit into any frequently occurring Main Code (Appendix-24), was assigned to the Potential Theme 'ER Culture was Lacking' (Appendix-25). Also, the whole Main

Code 'ER is Considered a Study Activity' in Phase-3, which had a significant number of extracts (19) and some Sub-Codes (Appendix-24), was also assigned to the potential theme 'ER Culture was Lacking' (Appendix-25). The envisaged narrative could describe how their reading usually focused on academic studies, and so perhaps, reading for pleasure was something new and strange, and so they could have been treating ER as just another study activity.

Appendix-25 also gives the number of extracts taken from each of the interviewees, for each of the Potential Themes. It shows that most were not unanimously voiced by the interviewees, and it shows that the distribution of extracts between them was never uniform.

Overall, during Phase-4, 11 Potential Themes were isolated, and these were arranged into 4 stages of narrative (Appendix-25).

At this point, four extracts are now given below as worked examples, to show how they were allocated into codes and then themes during Phase-2, Phase-3 and Phase-4. They give some illustration of the meticulous considerations made during this process.

Example-1: *'To be frank, reading is a problem, if in English or Arabic.'* (Transcript-Interviewee-04-p.11). This extract was first assigned the Initial Code 'ER Culture is Lacking' during Phase-2 (Appendix-23) because in describing even L1 reading as problematic, the participant was expressing he was not used to ER. Then during Phase-3 it was placed into the Main Code 'ER Culture is Lacking' & Sub-Code 'Abstract', which was used as an introduction to the Main Code (see Appendix-24), because this extract contained quite general wordings that was felt would serve well when opening up the narrative that would describe the students' lack of reading habit. Finally, it was placed into the Narrative 'Problems' & Potential Theme 'ER Culture is Lacking' during Phase-4 (see Appendix-25), as it clearly linked to an issue that could explain the modest reading and vocabulary gains noted during the programme (see Section-4.2 p.135).

Example-2: *'Both of my parent are lecturer in the university, so both of them they read a lot... but none of my family read! My brothers, no one.'* (Transcript-Interviewee-06-p.12). This extract was problematic because it seemed to indicate that ER was both present in one sense but lacking in another. Hence this was given two Initial Codes during Stage-2 – 'ER Culture is Present' because the participant clearly stated that his parents did read, and 'ER Culture is Lacking' because he clearly stated that the rest of his family did not (Appendix-23). This developed further during Stage-3 when it was placed into the Main Code 'ER Culture is Present' & Sub-Code 'Older People Read', noting the added detail that it was only some of the older generation that read, and also into the Main Code 'ER Culture is Lacking' & Sub-Code 'Abstract' (Appendix-24), because the description of the rest of the family not reading was a general statement with no reasons or explanation, and thus could serve as part of the introduction of that narrative. Finally during Stage-4 (Appendix-25) the first entry was placed into the Potential Theme 'Some ER Culture is Present' and this in turn was placed into the Narrative 'Open to ER if done differently', because it was felt that given the existence of some family members that do read, others could also pick up the habit if it was encouraged in an appropriate way. The second was placed into 'ER Culture is Lacking' within the Narrative 'Problems' given that it clearly linked to a potential reason why the vocabulary acquisition during the ER programme was limited (see Section-4.2 p.135).

Example-3: *'... I don't have time, so I can't read. You know the last semester, 32 hours.'* (Transcript-Interviewee-09-p.7 & Appendix-22). This was assigned the Initial Code 'Work Load at College' during Phase-2 (see Appendix-23), given that he was clearly expressing that his studies limited the time he had for ER. This was then placed into the Main Code 'Work Load at College' & Sub-Code 'Lessons were Too Many or Too Difficult' during Phase-3 (see Appendix-24), as he explicitly gave the reason as the large number of lessons he had every week. Finally it was put into the Narrative 'Problems' & Potential Theme 'The College Life Prevented ER' during Phase-4 (see Appendix-25), as this was again a potential reason that could explain the small gain in vocabulary during the ER programme.

Example-4: *‘When we entered this college we didn't take it seriously. We thought easy like xxxxxxxx college’* (Transcript-Interviewee-07-p.7). This extract was less straightforward because it did not easily fit into any single code, and so during Phase-2 (see Appendix-23) it was assigned to two: the Initial Code ‘New to Riyadh and the College’, referring to the first part of the extract that described the new situation he faced in the college, and the Initial Code ‘Teaching Style Differs from School & Family Experiences’, referring to the second part of the extract that described the very different situation in his previous college. Then during Phase-3 they were put into 2 Main Codes: the Main Code ‘New to Riyadh and the College’ and the Main Code ‘Teaching Style Differs from School & Family Experiences’ & Sub-Code ‘Lessons were Too Many or Too Difficult’ (see Appendix-24), which added the detail that the difference they faced in this college was that the lessons were more difficult than before. However it was finally placed during Phase-4 (see Appendix-25) into the Narrative ‘Problems’ & Potential Theme ‘The College Life Prevented ER’ alone, based on its Phase-3 Main Code ‘New to Riyadh and the College’ (see Appendix-24), as the Phase-4 Potential Theme ‘The College Life Prevented ER’ was judged to encapsulate both the Potential Themes under consideration during Phase-3.

These examples show some of the careful thought that was needed to process extracts that were not easy to place, and this was often the case given that the data was natural and unedited.

Phase-5 involved defining and naming themes in order to capture their essences and imports, with a focus on short, clear titles that will allow the reader to understand their meanings. Alongside this, it was necessary to further develop the narrative that started during Phase-4, in order to give the story that each theme told about itself, about how it interacted with other themes, and about how it related to the whole data (Braun & Clarke, 2006). The refined themes and how they mapped onto the Potential Themes isolated before in Phase-4 are given in Appendix-26.

Phase-6 involved producing the report, i.e. the physical writing up of the methodology of analysis (this Section-4.7), the results (below in Section-4.7.1 p.177), and the

discussion (see Chapter-5). Hence producing the report was considered, as recommended by Braun & Clarke (2006), to be an actual part of the analysis, and in fact ‘the final opportunity of data analysis’ (Vaismoradi et al., 2013:402). During this stage some further refinements were made, that were only realized to be necessary at this stage. For example, when producing the diagrams to summarise the final themes and their inter-relationships, a significant amount of time was needed to further refine the wordings of the themes (see Appendix-26), and to further refine the lay-out of themes within the diagrams, because it was thought, when looking at the write-up with a ‘fresh pair of eyes’ that more clarity was needed. Also, during this phase, it was decided to write the results (Section-4.7.1) as a narrative with quotes of extracts from the interviews, because it was thought that this method would present a more natural and compelling description of the participants and their complex feelings, hopes and lifestyles, and also ‘help the reader to comprehend the process, understanding, and motivation of the interviewer’ (Aronson, 1995). With this in mind, explicit reference to the ‘sub-themes’ was omitted, and instead they were referred to as examples of how the ‘themes’ were expressed, with associated quotes from the interview extracts (see below in Section-4.7.1).

4.7.1 Results

In this sub-section, ‘the most salient constellations of meanings present in the dataset’ are highlighted (Joffe, 2012). The previously described method of analysis yielded a number of ‘themes’ that offered insights into why the gains shown in the vocabulary tests were modest (see Section-4.2 p.135), and so directly addressed RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). However, more themes were found that went beyond RQ2, offering rich and detailed information about the students relating to their extrinsic motivation to study. Further themes were also found, and these were directly relevant to the future implementation of ER.

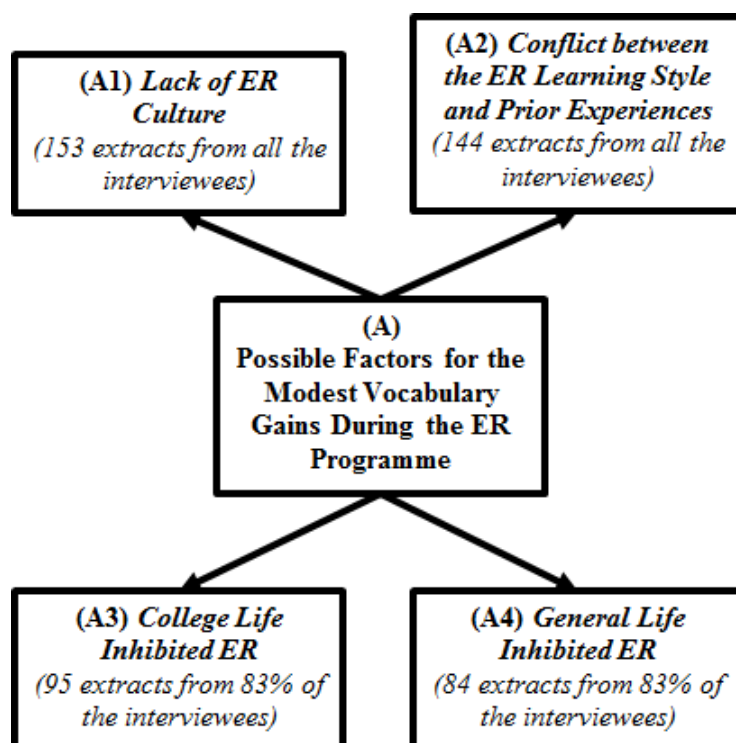
The last table of Appendix-26 gives the number of extracts taken from each of the interviewees for each of the final themes, noting, as defined previously when describing Phase-3, that a ‘theme’ should represent a patterned response or meaning that was

voiced by 75 per cent or more of the interviewees. It is clear from this that few of the final themes were unanimously voiced by all the interviewees, and furthermore, none of themes had a uniform distribution of extracts between the interviewees. For example, Theme-A1 was voiced by all interviewees but with a large spread of extracts between them: from 30 extracts (Interviewee-07) down to just 2 extracts (Interviewee-03). In addition, this spread of extracts differed from theme to theme. For example, both Theme-C3 and Theme-C4 were voiced by 92 per cent of the interviewees, but in Theme-C3 it was Interviewee-02 who did not voice the theme, while in Theme-C4 it was Interviewee-11. All of this variance was not surprising given the naturalistic nature of the interview data, and in fact it probably would have been unlikely to have seen uniform and unanimous results in such a situation.

The final themes are now presented in detail as a narrative with extracts from the transcribed interviews, accompanied with diagrams that summarize the main points.

For the primary aim of addressing RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), many themes related to **(A) Possible Factors for the Modest Vocabulary Gains during the ER Programme** (see Section-4.2 p.135), and these are summarised below in Figure-10.

Figure-10: Results of the In-Depth Interviews: (A) Possible Factors for the Modest Vocabulary Gains during the ER Programme



The arrows in Figure-10 are intended to show that all the themes (A1) to (A4) were possible factors that could have accounted for the modest vocabulary gains during the ER programme. Hence firstly, all the interviewees described in different ways (153 extracts) that there was a (A1) **Lack of ER Culture**:

‘In reality we are the culture of no reading... even in Arabic.’

(Transcript-Interviewee-06-p.12).

This could have caused some students to dislike reading

‘I don’t like reading... Most of them they don’t like reading’

(Transcript-Interviewee-06-p.4),

and this could have caused others to find it difficult

‘I think we are not used to have breaks to read. We are used to have breaks from reading [smiling]! Not to read!’

(Transcript-Interviewee-05-p.2).

Some described how they did not read for pleasure when they were young (e.g. Interviewee-08 (Transcript-p.10) said that he only started reading newspapers at college), and instead their reading was wholly connected to their academic studies:

‘They still have a problem with this, even in Arabic, because they didn’t read you know. Just for school and that’s it.’

(Transcript-Interviewee-05-p.9).

Even during the ER programme, reading was approached like an academic activity, by writing notes and memorizing new words:

‘I was taking the hard words to know them, then I translated them and then I recorded them.’

(Transcript-Interviewee-02-p.2).

Secondly, all the interviewees reported in different ways (144 extracts) that there was a **(A2) Conflict between the ER Learning Style and Prior Experiences**. They reported they had a strong motivation for grades

‘Yes and they will get like 1 mark, 1 point from 100, they have motivation.... Yes. Of course.’

(Transcript-Interviewee-05-p.15),

and some reported they would not read at all without them

‘Maybe I would not read if there are no marks.’

(Transcript-Interviewee-08-p.13).

Others stated that they expected the teacher to control their learning

‘We can’t see ourself. We can’t see what is a good book for us. The teacher should choose for us.’

(Transcript-Interviewee-08-p.2),

with the teacher being expected to actively check up on the students

‘... we will be careful if we know the teacher will ask us about the last lesson what we do.’

(Transcript-Interviewee-07-p.13),

and if not, it could be viewed negatively:

‘Many teachers, how can I say that, just let the student if they want to learn, this is your choice, but you don’t want that’s no problem. I just come to check present and teach you and then up to you.’

(Transcript-Interviewee-01-p.5).

Some expected a teaching style that was distant

‘The culture with a father and a son is a bit of far distance. So imagine with a teacher and a student. It much further.’

(Transcript-Interviewee-06-p.7),

or a style that was authoritarian

‘Yeh, I mean for many of my, for many of the students it’s like an order. Read, no discussion. Read, do your work. They think you know, they thought the teacher... like the teacher in high school.’

(Transcript-Interviewee-06-p.8),

and hence a free choice in their learning was not expected

‘Yes, for my specialisation, I was given a choice for my presentation. But this was the first time. Not before that college.’

(Transcript-Interviewee-08-p.4),

because their previous learning experiences were based on memorizing supplied facts

‘Er from I think, from primary school, from high school, from we were, we er were children, everything they taught us is fact, fact, fact. They didn’t try to have like a conversation. This is a culture thing.’

(Transcript-Interviewee-06-p.6).

In addition, most reported in different ways (95 extracts from 83 per cent of the interviewees) that **(A3) College Life Inhibited ER**, because of the heavy work load they had

‘...It about the number of classes...’ (Transcript-Interviewee-05-p.1),

all of which was not expected:

‘We thought easy like old college. Yes we take A+. 1 hour 2 hour before exam and good.’

(Transcript-Interviewee-08-p.7).

This was explicitly stated by some to be a reason for not being able to engage in much ER

‘I don’t have time, so I can’t read. You know the last semester, 32 hours.’

(Transcript-Interviewee-09-p.7 & Appendix-22),

and a reason for not liking ER:

‘Reading makes you remember this college. For me. Yes, really. Oh college, work, oh oh.... English, more English, more English. When you finish, I don’t like English class! You want some time for different, you’re tired.’

(Transcript-Interviewee-09-p.5 & Appendix-22).

An additional pressure was that almost all students were living away from their families

‘Most of us arrive er with no family...’

(Transcript-Interviewee-06-p.1),

which some reported was very difficult

‘You know the family is most important’

(Transcript-Interviewee-04-p.6),

and which others said made their studies even harder:

‘Because I hate the lonely life. When I come here to Riyadh, my performance go low. I feel home sick.’

(Transcript-Interviewee-01-p.6).

Some reported that English itself was a difficult language to learn

‘The English is difficult.’

(Transcript-Interviewee-12-p.3),

let alone as a language to use for ER:

‘Yes. Because it takes time to understand, it takes time to build the story in your mind, because you know you need to translate and build the idea and then it is confusing. Not easy in the beginning... You know when you start learning in another language, a new language for most of them, you get tired you know, very quickly.’

(Transcript-Interviewee-05-pp.3-4).

Finally, most reported in different ways (84 extracts from 83 per cent of the interviewees) that their **(A5) General Life** also **Inhibited ER**, citing many other pastimes that took precedence

‘Maybe I don’t like to spend time sit and reading.’

(Transcript-Interviewee-12-p.8).

Popular pastimes included watching films and soap operas

‘Every day I watch English films and soap operas....’

(Transcript-Interviewee-01-p.10),

watching documentaries

‘Documentary, it is very interesting to watch, because I already when I have nothing I just switch on the TV and watch the documentary. It is my favourite.’

(Transcript-Interviewee-05-p.13),

relaxing with friends

‘Sorry for I said that, but Saudi guys... Just entertainment, er, hanging out, go to the coffee shop, bowling [laughing]... like this.... But Saudi, the Gulf is very special in this.’

(Transcript-Interviewee-10-p.19),

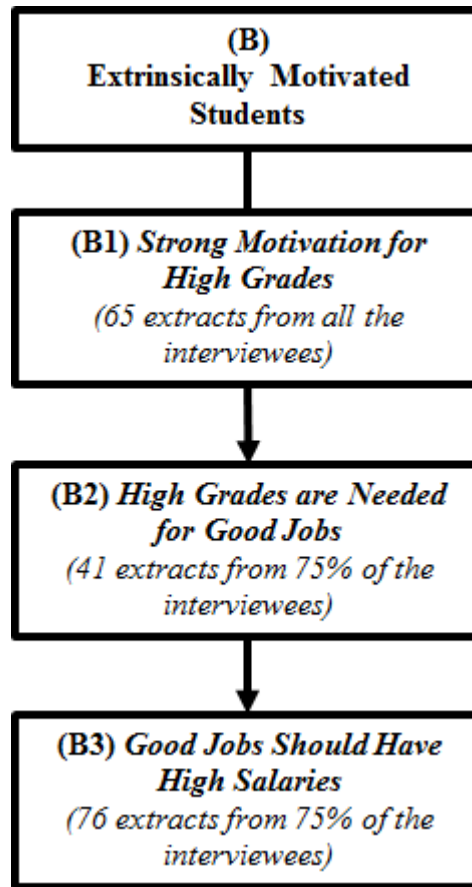
and hunting

‘Even you know Saudis they like to hunt you know... Yes, hunting, they are addicted [laughing].’

(Transcript-Interviewee-05-p.13).

The interviews also provided rich and detailed additional information about the students, namely that the participants were **(B) Extrinsically Motivated Students**, as summarised below in Figure-11.

Figure-11: Results of the In-Depth Interviews: (B) Extrinsically Motivated Students



The arrows in Figure-11 are intended to indicate that theme **(B1)** was caused by **(B2)** which in turn was qualified by **(B3)**. Hence all the interviewees expressed in different ways (65 extracts) a **(B1) Strong Motivation for High Grades**, which had been instilled in their previous educational experiences

‘From the primary school to the secondary school, the teacher always you must study for the exams, so all Saudi students have habit with this.’

(Transcript-Interviewee-01-p.5),

and in their general culture

‘It is something in our culture I think!... The most important is the certificate. If you have certificate, that’s it. Because your parents, they ask you have to get your certificate from that university and you have to get a job, and that’s it... This is the life in our parents’ imagination, because they were suffering in their childhood.’

(Transcript-Interviewee-05-p.4).

Most used different phrases (41 extracts from 75 per cent of the interviewees) to inform that this was because **(B2) High Grades are Needed for Good Jobs**

‘This is why the grades are important. In the future I want to find the job.’

(Transcript-Interviewee-04-p.7),

which was considered an important cultural responsibility

‘Maybe for er our culture. Man have to work... Our culture say this thing. If you don’t work, you are not a man.’

(Transcript-Interviewee-12-pp.4-5),

and a source of family pride

‘This the I think the main of our culture here. You want proud of self and your mother and father and get a good job.’

(Transcript-Interviewee-10-p.5).

Most added through different phrases (76 extracts from 75 per cent of the interviewees) that **(B3) Good Jobs Should Have High Salaries**, which made it even more important to get high grades. Some said they expected a high salary after graduation, especially if they were to work away from their home town

‘12 [thousand i.e. about 2000 GBP per month tax free]? Of course I would be happy.... But outside my city. Maybe 19 or 20. [i.e. about 3300 GBP per month tax free]’

(Transcript-Interviewee-10-p.16).

The good salary was needed to serve their parents

‘My father he give us a lot when I was children. Now is my turn to give him and help him.’

(Transcript-Interviewee-10-p.4),

and to serve their families

‘The first priority is how can I live and how can I spend money on my family. This is the first priority...’

(Transcript-Interviewee-12-p.13).

A very important task for the students was to get married soon after graduation

‘There is a lot [who get married 1 or 2 years after getting a job]. Already we have the students who is married.’

(Transcript-Interviewee-10-p.5),

but this required a lot of money

‘Yes, the wedding hall maybe 30000 [i.e. about 5000 GBP] is one night. For the dinner maybe also 30000, and 30000 for the wedding gift for the woman. And you need small, er, apartment... Yes, you need spend like this and more. This is simple marriage.’

(Transcript-Interviewee-09-p.10 & Appendix-22).

Some also reported other expenses that could only be realized with a well-paid job, like buying gifts

‘Sometimes I buy gift, sometimes I give them money. If he very close, maybe give him 1000 [i.e. about 170GBP]. If he just friend, give him 500.’

(Transcript-Interviewee-10-p.11),

buying a new car upon graduation

‘First job, er, I want er, Tahoe, Chevrolet Tahoe.... Of course new! In cash, it depends on the normal or extras... The normal 100, maybe 170000 [i.e. about 28000 GBP]. This is the normal.’

(Transcript-Interviewee-10-pp.10-11),

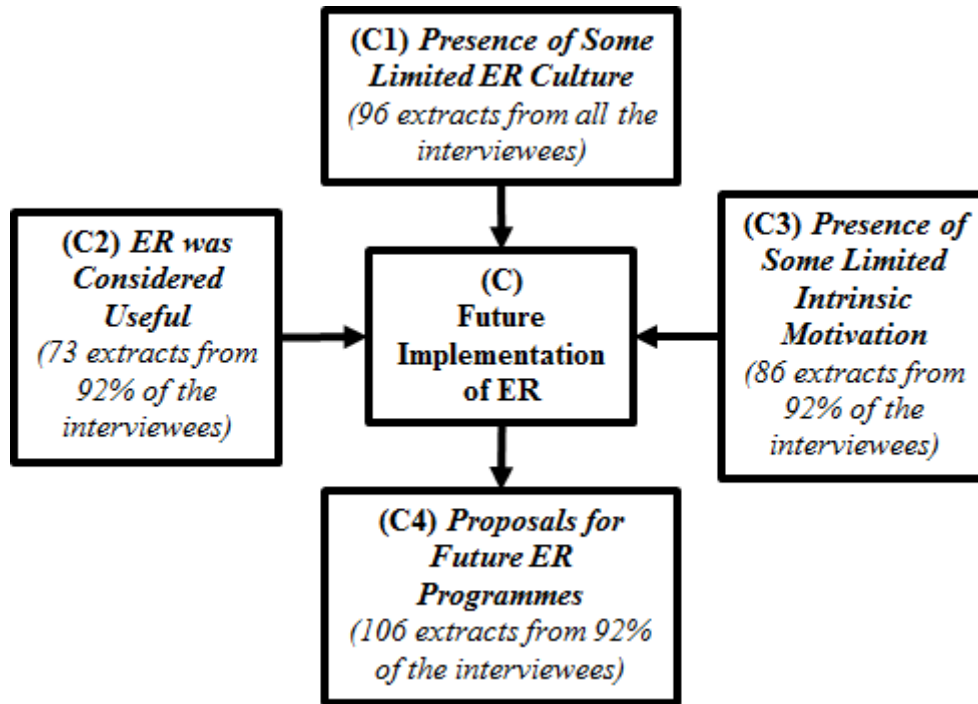
and setting up a business

‘How can be build or make a company in the beginning without money. A lot of businessman get a job in the government and then they save money and they quit from his job to start his private job. But first we need maybe work or get job.’

(Transcript-Interviewee-01-p.8).

Finally, other themes were directly relevant to the (C) **Future Implementation of ER**, as summarised in Figure-12 below:

Figure-12: Results of the In-Depth Interviews: (C) Future Implementation of ER



The arrows in Figure-12 are intended to indicate that themes (C1), (C2) & (C3) gave reason to believe that the ER programme could be more successful if administered differently, which thus led to the explicit suggestions of theme (C4). Firstly, all reported in a variety of ways (96 extracts) that there was a (C1) **Presence of Some Limited ER Culture**, but reading ‘conventional’ books seemed limited to some of the older generation:

‘No, no. Er 3 times bigger than this room full of books... Yeah he [i.e. his father] reads them, every day. Even my mother.’

(Transcript-Interviewee-06-p.12).

Some students reported that they did read ‘conventional’ material like newspapers

‘Yes, yes my brothers all read newspapers in Arabic.’

(Transcript-Interviewee-08-p.5),

and religious books

'I read real things like news and books related to Islam or something. Every day I am reading or every two days.'

(Transcript-Interviewee-07-p.5),

but the main ER they did was with their smart phones, with which they could spend many hours a day

'I was reading about a new application on iOS on iPhone, named Moment or something. That Moment application can access and analyse your action with the phone. That he can give you in the end of the day an information about how many er you know, how much time you spent on reading of Whatsapp, on Twitter, on er any website... So you know I was surprised I spent 7 hours of reading!'

(Transcript-Interviewee-06-pp.3-4).

Hence social media was reported to be popular

'Saudis now they most use the Twitter. And if you go on any news for Saudi guys you can find they are make follow for er the Arab News, for Saudi news. About the update, what's the news. What it will be in the future. What the Ministry of Labour says...'

(Transcript-Interviewee-11-p.6),

as well as watching videos with subtitles

'I'm watching an English film every 2 or 3 days, not including the English soap operas, but with Arabic subtitles.'

(Transcript-Interviewee-02-p.10).

In addition, it was clear through many different phrases (73 extracts from 92 per cent of the interviewees), that **(C2) ER was Considered Useful:**

'I believe it is important, but I didn't read... I don't know why I don't like it. Just. I don't know the reason.'

(Transcript-Interviewee-12-p.8).

They acknowledged it was useful for general learning

'If you read you learn something new.'

(Transcript-Interviewee-12-p.7),

exercising the mind

'... my father always tell me the mind need eat the mind reading... Each time take one hour or two hour for reading ... He tell this.'

(Transcript-Interviewee-09-p.7 & Appendix-22),

assisting in their studies

'Reading from the book, but when I find is it advantage for me maybe next semester. The semester because in our specialization there is a lot of books and we have to to read read a lot of books.'

(Transcript-Interviewee-11-p.1),

and for improving their English language skills

'Main reason I am reading in class is to improve my writing and grammar.'

(Transcript-Interviewee-07-p.13).

In addition, some also reported that the ER programme improved their reading

'Confidence, spelling and speed. Because before, as I said, one by one like, er, like a children, word by word. But now no, I can read the line very fast, and I understand. Fast and understand at the same time.'

(Transcript-Interviewee-10-p.19),

and their vocabulary

'For me, ar its was ar very useful. For me. Because some vocabulary I know what's the meaning. I know when I use this er word, but I never see the spelling.'

(Transcript-Interviewee-12-p.3).

Furthermore, most expressed the **(C3) Presence of Some Limited Intrinsic Motivation** in different ways (86 extracts from 92 per cent of the interviewees). For example, the grades were occasionally not important

'I say I don't care about the mark, just correct it er way and I remember he gave me 83%, but he change everything.'

(Transcript-Interviewee-06-p.10),

and even simply pleasing the teacher was occasionally more important

'Even if no grades, I like tests, because the teacher will ask me the next day what you read.'

(Transcript-Interviewee-07-p.14).

The participants mentioned different things they did to improve, without being asked to do so, like trying to surround themselves with more English when using the internet

‘Sometime the Twitter and social programmes, this is where you find the read English. Download the each word and sometimes the story in English. I want to practice.’

(Transcript-Interviewee-04-p.10),

and when using their computers in general

‘Also my computer, my laptop it’s already in English.’

(Transcript-Interviewee-03-p.2).

Vocabulary learning featured many times as an important self-study that some of them and their friends willingly undertook, even just when watching films

‘I found another problem with the movies, if you find a phrase and look at each word separately you will find another meaning. For example in one sketch I heard the characters say ‘give him company’. If we separate it it means ‘give him a company / business’. Doesn’t make any sense. But the meaning of this phrase ‘spend some time with him’.’

(Transcript-Interviewee-01-p.11),

or television

‘I remember my cousin, he is now engineer in Qatar, he had a problem with English words, it was before smart phones. So he has like an electronic dictionary so he watch an episode, so he stop and translating every word...’

(Transcript-Interviewee-06-p.10).

With this background, most gave many explicit **(C4) Proposals for Future ER Programmes** (106 extracts from 92 per cent of the interviewees), pointing out that it was the way the ER was presented that needed to change

‘Reading is benefit but the method has to be more fun.’

Transcript-Interviewee-07-p.11).

Many of these proposals were clearly linked to previous themes. For example, some interviewees recommended reading activities that contributed to the final grade

‘I think yes, yes. I think that would be great if connect to read. Fantastic. Because I already saw my friends when they have low marks, they ask the teacher you know, give me a homework, give me a task, anything, I will do it to get extra marks. And it works, because now they have a challenge, they have a challenge, because they want to pass to get certificate. If you do it like this, I think it will be better. Yes, they will read.’

(Transcript-Interviewee-05-p.7),

especially vocabulary learning activities

‘Er, in the past some teacher er he give us idea every day or every week give us maybe 8 to 15 words... Yah. And save it, er memorise it, and in the end of the weeks they give us like a test. This word, what it means, this vocabulary what it means, like this. This is good idea, this works with me.’

(Transcript-Interviewee-10-p.9).

Many other activities were also suggested, such as discussions

‘... we start reading for maybe 30 minutes, and then have break just to discuss if they are interesting, if they are not. Just break for 10 minutes.’

(Transcript-Interviewee-05-p.3),

presentations

‘I feel I will be excited to er explain the story or be like a play. ... I think it’s more challenge. It’s better. I thinks.’

(Transcript-Interviewee-10-p.2),

and read-aloud summaries

‘...and write 1 paragraph and they have to read it for students, in front of the students, yes... So, and it will be more interesting. They have a challenge.’

(Transcript-Interviewee-05-p.14).

Other (C4) **Proposals for Future ER Programmes** related to the Sustained Silent Reading sessions. Some participants said they should be allowed to read the internet

‘More interesting to allow iPads and let students read websites... Attractive. Change from the normal way of study.’

(Transcript-Interviewee-08-p.8),

or watch videos with English subtitles

‘Of course it will be more interesting.... Because you know most of us, most of us are used to watch movies and so on, so it is interesting to watch something... Yes, of course...’

(Transcript-Interviewee-05-p.13),

or that they should be guided to the books they should read

‘Again must guide the student to a level that's OK for him. The stories are interesting but the level is too hard.’

(Transcript-Interviewee-07-p.12).

Finally, some suggested using technology to make filling in the Reading Diary Sheets quicker, less laborious and more accurate

‘Another thing, if I was a teacher and I wanted to assist the students for writing er a sheet about that, I'd probably use a website or something. There is a new technology named Google Forms...’ and ‘I was reading about a new application on iOS on iPhone, named Moment or something... That he can give you in the end of the day an information about how many er you know, how much time you spent on reading of Whatsapp, on Twitter, on er any website... So you know I was surprised I spent 7 hours of reading!?...’

(Transcript-Interviewee-06-pp.3-4).

4.7.2 **Summary**

In summary of all the above, the interview data gave insights into why the gains shown in the vocabulary tests were modest (see Section-4.2 p.135), and so directly addressed RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36). Key factors voiced by most (at least 75 per cent - see above in Section-4.7.1 p.177) of the interviewees were that they were not used to reading for pleasure; the

skills needed for ER were very different to those needed during their previous educational and life experiences; and their lives, both inside and outside of college, preoccupied them from reading for pleasure.

In addition, more themes were found that went beyond RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), offering rich and detailed information about the students. In particular, they showed a strong motivation to get high grades, because in their society high grades were the most important factor to get well-paid jobs, which they needed to support the many expenses that were important to them.

Finally, further themes emerged that were directly relevant to the future implementation of ER. Hence, the participants reported some limited interest in reading for pleasure, the presence of some limited intrinsic motivation when studying, and the acknowledgement that ER was useful, all of which gave reason to believe that if the ER programme was administered differently, it could be more successful. Following this, the students themselves gave many specific suggestions, including testing the reading; giving grades for the reading; adding vocabulary activities, discussions, presentations & summaries to the reading; and making more use of technology such as tablets to encourage reading the internet and watching films with subtitles. A detailed discussion of these results is presented in Chapter-5 (Discussion).

CHAPTER-5

DISCUSSION

In the previous chapter, a detailed description of the Results was given. In this chapter, a detailed Discussion of those Results is now presented by meticulously extracting and debating the many important issues that arose.

5.1 DISCUSSION OF RESULTS RELEVANT TO RESEARCH QUESTION 1

As stated before in Section-2.8 p.36 the first research question of this study was:

Research Question 1 What is the impact of an extensive reading (ER)
(RQ1) programme of a relatively short duration, on the
 vocabulary acquisition of male Saudi university students?

For this, pretests and posttests of vocabulary knowledge were administered to both the experimental and control groups to gauge vocabulary gain, as detailed in Section-3.2 p.54 & Section-3.4 p.72.

The results from these tests are now summarised, scrutinized and discussed in order to offer an informed response to RQ1 that is presented at the end in the summary Section 5.1.1 p.195.

Section-4.2.7, Figure-6 p.138 & Appendix-18 detailed the results of the vocabulary tests, and they showed that there were few statistically significant vocabulary gains after the ER programme (13.1 weeks – see Section-3.2.5 p.66), and that these were in the more frequent word levels. In general, the gains for Test-4 (Meaning through multiple choice questions (MCQs)), which was based on the Vocabulary Levels Test of Schmitt et al. (2001) as detailed in Section-3.4.5, were more than the other 3 tests that were based on the Vocabulary Size Test of Nation (2009) as detailed in Section-3.4.4 p.79. For these, Test-2 (Open Translation) had better gains than Test-1 (Spelling through

MCQs). In addition, the results involving the different levels of certainty (i.e. Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone) showed that none of these three showed consistently better statistically significant gains over the others.

The first natural point of discussion arises from comparing these results with those of the other studies that were described in Chapter-2 (Literature Review), even though it should be noted that this study had several distinctive aspects that aimed to address the limitations of those previously surveyed studies (Section-2.8.2 p.38), making unqualified comparisons impossible. Hence, in the study of Al-Homoud & Schmitt (2009:396), both the ER and IR groups showed similarly slight vocabulary gains, being broadly comparable to this study's results. This could be a reflection of the part similarity of students in both studies, given that both sets of participants were young-adult male Saudis studying during their university preparatory courses.

However, in terms of the different aspects of word knowledge tested, the study of Pellicer-Sanchez & Schmitt (2010:41) showed that their results for Spelling through MCQs were better than those for Meaning through Open Response, and other studies such as those of Pigada & Schmitt (2006:8) and Waring & Takaki (2003:141) also found greater vocabulary gain for Spelling than for Meaning through translation or open response. These contrasted this study's results which found that gains for Open Translation (Test-2) were more than for Spelling (Test-1), even though the gains for the receptive tests were expected to be better than those for the productive test, as found by Ishii & Schmitt (2009:12) and as first mentioned in Section-3.4.4.2 p.84.

The poorer gains found in this study for Spelling (Test-1) could have arisen from this aspect of vocabulary acquisition being particularly problematic for the Arabic speaking participants, compared to the non-Arabic speaking participants of other studies. Arabic words are almost entirely pronounced in accordance to their spellings, to the extent that it is not normal to test Arabic speaking children on their spelling. However English spellings are notoriously difficult because of their 'illogicality... [and] ...arbitrary nature' (Moody, 1974:318) when compared to what could be plausible based on

pronunciation alone, as first mentioned in Section-3.4.4.1 p.82. In addition, the Arabic script is completely different to that of English, and is written in the opposite direction. All of these factors could have contributed to spelling being more challenging for the students of this study's context, compared to those in other studies.

Furthermore, while developing their vocabulary levels test, Schmitt et al. (2001:69) found that speakers of non-Romance languages found the Academic-word level difficult. This was supported in this study where there was less vocabulary gain in this section than in the 2000-word and 3000-word sections of Test-4 (Meaning through MCQs) (based on the Vocabulary Levels Test, Schmitt et al., 2001, as detailed in Section-3.4.5), but at the same time there was a greater improvement in Certainty in the Academic-word list than the other two. This could have occurred because the students of this study were at the start of their university studies and so were being exposed to lots of new academic vocabulary, which was slowly becoming more familiar. However, given that the study spanned only the first semester of their studies, they had not gained enough exposure of these words to respond correctly in the posttests, but at least they were able to report by that time that they were more confident with their attempts.

The results of this study's vocabulary tests (Section-4.2.7, Figure-6 p.138 & Appendix-18) also showed greater statistically significant vocabulary gains for the Experimental cohort than the Control, but the difference was not large in terms of absolute numbers of words, and in a few cases the Control actually showed slightly more vocabulary gain. These modest gains could have been because the learners were beginners, noting that more advanced learners can pick up more vocabulary than beginners (Waring & Nation, 2004:103), or it could simply be the case that more than just one semester of ER is needed to see better gains, as recommended by Kirchhoff (2013:193) and Clark & Ishida (2005:235), with longer programmes (Asraf & Ahmad, 2003:99) and more in-class silent reading sessions being needed (Mermelstein, 2015:194). Such longer programmes could also allow for more differentiation to be seen in terms of Certainty, which was one of the aspects of partial vocabulary acquisition addressed in this study (see Section-3.4.7 p.90).

In addition, perhaps some explicit vocabulary learning is also required, as suggested by Clark & Ishida (2005:236-237), but this would contradict the principles of pure ER. Furthermore, perhaps the choice of some students to read unsimplified materials may not have provided for them the required repetition of words that are thought needed for incidental vocabulary acquisition, noting that some recommend more than 10 exposures of a word for such uptake (Pellicer-Sanchez & Schmitt, 2010:42), while others recommend more than 20 (Pigada & Schmitt, 2006:19). It would seem then, that it could be beneficial to give at least as much encouragement to read GRs, 'where repetition and recycling are a basic principle' (Pellicer-Sanchez & Schmitt, 2010:44), even though sometimes GRs have been found to not have enough word repetitions (*ibid.*, p.47) as first pointed out in Section-2.7.1 p.20.

Another interesting point of discussion is the role of teacher enthusiasm in the success of ER programmes in general, and in this programme in particular. Mermelstein (2015:194) reported the view that actively encouraging reading is a factor for successful ER, and Asraf & Ahmad (2003:99) asserted that in order to motivate the students to read, the teachers themselves must have and show positive attitudes to ER. Littlejohn (1985:257) went further to suggest this applies to any theory held by the teacher: if a teacher holds a particular theory, the students will respond favourably when this theory is used with them, and they will become more motivated in the associated tasks. However, as described in Section-3.2.4 p.65, the researcher showed great enthusiasm when administering the ER Programme, actively encouraging the participants to read, and actively trying to present reading in an enjoyable way. Yet with all of this, the vocabulary gains (Section-4.2.7 p.147) were modest and not consistently better in the Experimental Cohort when compared to that of the Control Cohort, showing that the teacher's positive attitude was not sufficient in terms of enhancing incidental vocabulary learning through ER. This again could be because of the particular nature of the students involved, given that they read little for pleasure and so they needed much more motivation than the students of other studies who were already used to a higher level of such reading.

It should also be noted that this part of the study was experimental in nature, aiming to see ‘if one variable influences another by holding other factors constant and varying the treatment given to two groups’ (Hyland, 2002:170). This was a deductive approach that ‘begins with a hypothesis or theory and then searches for evidence either to support or refute that’ (Nunan, 1992:13). In this context, the aim was to see if the ER Programme in the Experimental Cohort would be associated with enhanced vocabulary acquisition when compared to that of the Control Cohort. However, even though an ‘increased reliance on experimental research’ (Kluge, 1997:3) has been reported within the research community, it should be remembered that ‘classrooms are not laboratories’ (Hyland, 2002:171) and as such, it is almost impossible to hold all the other variables constant within the experimental and control groups. Hence it has been asserted that experimental research ‘cannot really be generalized to all teaching contexts’ (Kluge, 1997:3), and so caution should be exercised when trying to do so in an unqualified manner with these results, even though the ‘desire to make the strongest claims we can is a perfectly natural one’ (Richards, 2003:4). Instead, these results should be viewed within its context and only applied to other contexts if they can be seen to be similar.

5.1.1 Summary: Response to Research Question 1

Hence to summarise the above, it can be suggested in response to RQ1 that the ER programme of a relatively short duration that allowed an open choice of reading materials had a modest impact on the vocabulary acquisition of male Saudi university students.

When comparing this with the results of other studies, it would seem that the different characteristics of the participants could explain the smaller vocabulary gains, and it could also explain why the gains for receptive knowledge of Spelling were less than those for productive knowledge of Meaning. However, given the challenging context of the study (see Section-2.8 p.36), some significant vocabulary gains were still found and these were generally slightly higher for the Experimental Cohort than the Control, in terms of both the students’ correct answers (especially words from the more frequent word levels) and the students’ confidence when attempting answers (for example the

Academic words). This could indicate that ER still has a role to play in aiding vocabulary acquisition in these students. To achieve this though, some established aspects of ER may need to be strengthened, such as having a much longer programme and having more in-class silent reading sessions, and further amendments that are not typically part of pure ER and that may even contradict its principles are also possibly needed, such as explicit vocabulary learning and teacher selected reading material. It should also be noted that the great enthusiasm of the teacher (in this case the researcher) in presenting ER (see Section-3.2.4 p.65) did not appear to be enough to motivate the students' reading, which again could have stemmed from the lower levels of pleasure reading that they were used to when compared to those in other studies.

Given that all these important points of discussion seem to revolve around the distinct nature of the students and their circumstances, it should be noted that the results of this more experimental side of the study should be considered in light of its specific context, and the temptation to generalise them to situations beyond that should be resisted unless they can be demonstrated to be comparable.

5.2 DISCUSSION OF RESULTS RELEVANT TO RESEARCH QUESTION 2

The second research question of this study (Section-2.8 p.36) was:

Research Question 2 What characteristics of the students, in terms of the nature
(RQ2) of their reading, their other English learning, and their
 culture, may also be associated with any observed
 vocabulary acquisition?

For this, Reading Diaries were completed, a Survey of Other Sources of English Learning was administered, and In-Depth Interviews were conducted (Section-3.6 p.103, Section-3.7 p.111 and Section-3.9 p.118), all to gauge factors that could be associated with the modest vocabulary gains found (Section-4.2.7 p.147) during the ER Programme (Section-3.2 p.54).

This part of the study was more inductive in nature, seeking to ‘derive general principles, theories, or ‘truths’ (Nunan, 1992:13) from the naturally occurring data. This was a more ‘data-driven..., ‘bottom-up’, or ‘heuristic’ (McDonough & McDonough, 1997:79) approach that complemented the more deductive approach adopted to address RQ1 (Section-3.1.3 p.50 & Section-5.1 p.195), making the study more acceptable to all researchers in the field, regardless of their philosophical backgrounds.

The results of each of the relevant data collection instruments are now abridged examined and debated in order to propose an informed answer to RQ2 that is given at the end in the summary Section-5.2.4 p.212.

5.2.1 Discussion: Reading Diaries

The results for these were detailed in Section-4.4.1 p.154 & Appendix-20 and are summarised here. Firstly, an average 8.84 hours/participant/week and 2.46 hours/participant/week of English and Arabic reading was recorded respectively. The amount of English reading can be described as very modest, given that the participants were studying their bachelor degrees in English, and yet it was still much more than

their L1 reading. This suggests that reading in itself was a skill they were not used to, and so reading in the foreign language of English could be even more challenging. Yamashita (2004:13-14) reported that L1 reading attitudes do transfer to those for L2, which prompted the recommendation to encourage L1 reading in order to promote L2 reading. This could be particularly relevant to this context where perhaps an Arabic ER programme before or alongside the English ER programme could be beneficial, even though such a bold idea would require further resources and support that may not be available in many institutions.

Also a large proportion of the English reading material (see Section-4.4.1 p.154 & Appendix-20) was related to their college work (average 3.79 hours/participant/week), with reading college textbooks being the main part of this (average 2.99 hours/participant/week). Hence the text book was an important part of the students' college life, which was not dissimilar to their previous educational experiences, and which is not dissimilar to the situation elsewhere in the world where textbooks are the 'core of many teaching programmes' (Brown, 2009:240), providing security to the students because they act as a road map that allows them to know what to expect in the course (Graves, 2000). This could be capitalized upon by using textbooks that explicitly promote ER, as recommended by Brown (2009:240) who stated that this can give 'credibility and legitimacy' to ER in the eyes of such students, and hence this can encourage them to read more.

Furthermore, reading internet text (see Section-4.4.1 & Appendix-20) was the second biggest amount of English the students read (average 1.85 hours/participant/week) and it was the biggest amount of Arabic read (average 1.05 hours/participant/week). Given the popularity of browsing the internet, this could be an important source of ER that can be actively encouraged, especially for ER outside the classroom, as envisaged by Day (2015:297). The next biggest amount of Arabic reading was that of Arabic subtitles in English films (average 0.98 hours/participant/week), which again could point to an entertaining source of ER, which when done with English subtitles has been considered one of the best ways to learn vocabulary (Meara, 2005:5).

It was also clear from these results (see Section-4.4.1 p.154 & Appendix-20) that most of the students' reading was not for pleasure, but instead for their obligatory academic work. This is further reflected in other items of the Reading Diaries, such as the results for Item-11 (How Essential the Reading was) (Section-4.4.1 & Appendix-20), which showed that most (about two-thirds) of the English reading was rated 4/5-5/5 (with 5/5 being 'really essential'). This echoed what Al-Homoud & Schmitt (2009:387) reported from another study of Saudi students that they had 'very low levels of pleasure reading in English', which again could indicate that a particular challenge in this context is to build the habit of reading in the first place, before being able to expect large scale ER in English to take place. Hence, their lack of an existing habit to read for pleasure could be an important factor that led to the modest vocabulary gains (Section-4.2.7 p.147) during the ER Programme (Section-3.2 p.54).

Although the results above seem to strongly suggest that the students were extrinsically motivated to read, other items of the reading diaries (Section-4.4.1 p.154 & Appendix-20) seemed to indicate different attitudes. Most of the English reading was done 4, 5 or 10 times/week, perhaps indicating that some were reading every week day only for their studies, but perhaps also indicating that others were reading much more, perhaps for pleasure. Also, about two-thirds of the English reading was rated 4/5-5/5 for Ease (with 5/5 being 'very easy'), and about two-thirds was rated 4/5-5/5 for Enjoyment (with 5/5 being 'very enjoyable'). This could suggest that despite the new experience of increased reading of English at college, the students were still very positive about it, which in turn could show that ER can still have an important role in improving their reading, vocabulary and overall proficiency, once they have become perhaps more accustomed to the skill of reading itself.

In summary, the above discussion of the results of the Reading Diaries (see Section-4.4.1 & Appendix-20) yielded many possible factors that could shed light on the modest vocabulary gains (Section-4.2.7) found after the ER course (Section-3.2), and it also fielded many possible modifications that could lead to better gains in future.

The amount of English reading recording was low and the amount of Arabic reading was lower still, indicating that they were unaccustomed to reading even in L1, which could lead to a bold suggestion to implement an Arabic ER course before and / or alongside the English ER course. Most of their reading was related to their college work using textbooks, which could be used in an advantageous way if textbooks could be found that actively encourage ER, making ER seem more formal and important to them. Reading the internet was popular, as was reading (Arabic) subtitles in English films, and these were in stark contrast to the bulk of their reading which they reported they were forced to do for their studies. Hence given the enjoyable nature of the internet and subtitled videos, these could be excellent sources of reading that could be encouraged more, although for the latter it would seem more useful to move the students on to reading English subtitles when watching English videos. Finally, most students did report that they enjoyed much of their reading and found much of it easy, which could indicate that ER can become more useful to them once they perhaps become more familiar with it.

5.2.2 Discussion: Survey of Exposure to English Learning

The results for these are detailed in Section-4.5.1 p.159 & Appendix-21, and they provide a profile of the participants' exposure to English outside their current university.

Item-1 (Age) and Item-2 (Formal Study of English Outside the Current College) showed the participants were predominantly in their early twenties, coming straight from college, which was the first time most had been taught English in English with native English speaking teachers, and almost all of their academic learning was in Saudi Arabia, where English is not used much outside the formal education setting. As touched on before (Section-3.7.1 p.112), age is an important psychological factor in language learning (Benson, 2004:10), and it could be related to background knowledge, which is thought to aid text comprehension (Nation & Coady, 1988:102-104; Schmitt et al., 2011:38), especially when the text contains many exophoric references (Cutting, 2008:8). This, coupled with their relatively short exposure to English learning with

English-speaking teachers, could indicate that they found reading English a relatively new and perhaps uncomfortable experience, which could have inhibited better vocabulary uptake.

Item-3 (External English Exams) showed that few had taken external exams, none had made any special preparation for them, and their scores were the equivalent to C-D grades in the Saudi Arabian educational system. This can indicate a modest proficiency in their tested English, and also a lack of development that could have taken place if more preparation was made for such exams, all of which could have contributed to modest vocabulary gains (Section-4.2.7 p.147) during the ER Programme (Section-3.2 p.54).

Item-4 (Travel Abroad) indicated that most had travelled for tourism, but few had travelled outside the Arab world, resulting in little English being used during such trips. The few that had used a lot of English were the older participants aged over 30, and they had been sent for training. Hence, for most, travel abroad had not greatly enhanced their English proficiency, and so again their lower-level ability in English remained and thus could have contributed to the slight vocabulary uptake.

Item-5 to Item-8, pertaining to the frequency and amount of English watched and listened to, showed many participants watch English videos every day, but mainly American movies, and often with Arabic subtitles. This could indicate that English did not prevent them from enjoying an activity, and so videos could be used to improve their English in an entertaining way, which could then positively affect their reading. Also, as noted in Section-5.2.1 p.201, if the students can be guided to watch English videos with English subtitles, this could improve their reading and help them learn vocabulary (Meara, 2005:5).

Finally Item-9 (Work Experience) showed that few had any work experience, and it was the older participants who had used the most English while working in the civil service before coming to the university. It could thus be proposed that formal English language teaching should continue after they leave the college, maybe on a part time basis, to go

hand in hand with the English they are using at work. Hence, an ER programme that is implemented during this time could be beneficial, given that the poor vocabulary knowledge of beginners inhibits their reading ability (Al-Homoud & Schmitt, 2009:387), and so continuing the reading while their proficiency gradually improves at work, could see better results.

In summary, the above discussion of the results of the Survey of Exposure to English Learning (see Section-4.5.1 p.159 & Appendix-21) yielded many possible factors that could shed light on the modest vocabulary gains (Section-4.2.7 p.147) found after the ER course (Section-3.2 p.54), and it also raised considerations for future courses.

The surveys recorded that the students were young adults, who had learned little English with English speaking teachers, who used little English outside their English lessons, who had typically not prepared themselves for any external English exam, and who had used little English during their travels. All of this showed their basic proficiency level that could have been a major reason for the modest vocabulary gains during the ER course. However, the survey also recorded that they frequently watched English videos with subtitles, just as was also recorded by the Reading Diaries (see Section-4.4.1 p.154 & Appendix-20), which again could be used as an enjoyable medium that can encourage more reading. Finally, a few participants recorded that they had used a great deal of English at work, which could justify the proposal to continue ER during the students' work placements, thus giving them input in terms of both reading (from the ER) and listening (from interaction with colleagues) that could enhance their overall proficiency dramatically.

5.2.3 Discussion: In-Depth Interviews

The results for these are detailed in Section-4.7.1 p.177, and they were of particular importance because, although the Reading Diaries and the Survey of Exposure to English Learning gave information that could be used to postulate reasons for the slight vocabulary gain (Section-4.2.7 p.147) during the ER Programme (Section-3.2 p.54), the In-Depth Interviews could present reasons directly from the students' mouths.

The most prominent theme expressed was the 'Lack of ER Culture' (Section-4.7.1 p.177 & Figure-10). The participants reported that they were not used to reading; that they did not like it; that they found it burdensome; and that they considered it an academic task. Waring & Nation (2004:103) reported that beginners need more encounters to learn a word than more proficient learners, and so in this study, the simple fact that they were not used to reading could be a big factor why not much vocabulary acquisition was seen during the ER Programme. Also, Yamashita (2015:169) considered large amounts of reading to be 'the essence of ER' and this can only be achieved when the reading is enjoyed (*ibid.*, p.173). However, the students in the study reported that they did not enjoy reading, so large amounts of reading were unlikely to follow, which in turn could yield limited vocabulary gains. Many of the previously surveyed studies reported positive attitudes to ER (Al-Homoud & Schmitt, 2009:399; Macalister, 2008:254), but this was not the reported in this study, adding to the few studies that reported less than positive attitudes to ER (Ro, 2013:216).

Also, the enthusiasm shown by the teacher when presenting the ER Programme (see Section-3.2.4 p.65) did not seem to motivate the students enough in their reading, even though positive attitudes of teachers have been considered 'crucial' (Littlejohn, 1985:260) for the success of the teaching methods they use. Furthermore, Ro (2013:215) listed the nature of the materials, the attitudes towards L2 reading, the L2 reading ability and the socio-cultural environment to be variables that affect the motivation to read in a second language, and given that attitudes to L2 reading are often affected by L1 reading attitudes (Yamashita, 2015:173), it can be seen in this study that three of these four variables are directly influenced by the lack of ER culture the interviewees reported.

Another reported theme was a 'Conflict between the ER Learning Style and Prior Experiences' (Section-4.7.1 p.177 & Figure-10). Hence the interviewees reported that their studying had been based on textbooks and the memorization of facts; that they expected the teacher to choose their learning materials; and that they expected the teacher to check up on them. Hence they were surprised that the ER Programme did not

have any of these features, echoing a similar situation reported by Kirchhoff (2013:193) in other countries when students are faced with ER, which could suggest that for these contexts, the ER programmes should be modified.

Brown (2009:240) recommended for the context where textbooks are heavily relied upon, that textbooks themselves should be used that explicitly promote ER. In a similar vein, Macalister (2008:255) suggested that at least some texts should be chosen by the teacher for the whole class, even though that contradicts the ER principle of giving learners the freedom to choose what they read. He argued that such an approach could help to formalize the ER programme, make it more acceptable to learners and administrators, and if these texts are chosen with other topics and subjects in mind, it could help to integrate other parts of curriculum. Furthermore Sivell (1980:52-54) proposed that where memorization is part of the students' culture, then this can be harnessed, given that memorization does at least show effort on the part of the students, and given that it does give the students a sense of confidence. In this context, the students can be actively encouraged to memorise vocabulary they have met during the ER programme, which again is not a feature of ER, but which could improve their attitudes towards ER and hence eventually lead them to read purely for pleasure.

The interviewees also reported that 'College Life Inhibited ER' (Section-4.7.1 p.177 & Figure-10). They informed that they had too many lessons to have time to read for pleasure; that most of them missed their family and so they could not focus on their studies well; and that English itself was difficult to learn, let alone to study all their other subjects in.

Many researchers have also noted the problem students have in finding time for pleasure reading (De Burgh-Hirabe & Feryok, 2013:90; Kirchhoff, 2013:193), with Mermelstein (2015:194) reporting the view that this is one of the factors that affects the success of any ER programme. In order to mitigate this, De Burgh-Hirabe & Feryok (2013:90) insisted that reading outside the class should to be well monitored or that more supervised in-class reading should be done, because the pure ER characteristic of voluntary reading would be overshadowed by the students' obligatory academic work.

Ro (2013:227) gave further suggestions such as actively explaining the value and usefulness of ER, and such as providing a dedicated reading location like a library where the students can shut out the rest of their concerns and instead just relax and read.

Researchers have also acknowledged the ‘twofold challenge of learning both the language itself, and simultaneously using it to learn the subject matter’ (Dufficy, 2004:242), and hence ‘students are often under tremendous pressure’ in EAP contexts (Komiyama, 2013:150). Given that English itself was reported to be a difficult language to learn, there could be what Ro (2013:214) described as ‘foreign language ‘reading’ anxiety’ that is thought to be caused by differences in the cultures and writing systems of L1 & L2, which in this context (Arabic and English respectively) could be very significant, given the different cultures and the different written scripts.

Similarly, the interviewees also reported that their ‘General Life Inhibited ER’ (Section-4.7.1 p.177 & Figure-10) with their many pastimes that took precedence, such as watching films, watching soap operas, watching documentaries, relaxing with friends, and even hunting. De Burgh-Hirabe & Feryok (2013:89) echoed this problem, noting that some students in their study couldn’t continue ER even though they reported they enjoyed the material, simply because they couldn’t organize themselves to find the time for ER amongst their other pastimes. Ro (2013:227) added that adult learners simply ‘lack the time to read; they are busy with their lives’, and Kirchhoff (2013:193) also noted that working adults have both work and a social life that can relegate ER once they have left the classroom.

However, some of these pastimes could present opportunities for ER. The internet was reported to be the main source of their films and programmes, and so this could be actively used as a major source of reading too, as envisaged by Day (2015:297) to be an enjoyable medium in this modern age for reading beyond the class. Coupled with this, videos with subtitles have been recommended as a ‘fun’ way to learn vocabulary (Meara, 2005:5), so these could also be a successful source of reading. However, it may be more challenging to find ways to encourage reading through social pastimes like

relaxing with friends, or through active pastimes like sports and hunting, given that reading can be considered both unsocial and inactive.

The results of the In-Depth Interviews (Section-4.7.1 p.177 & Figure-11) also yielded another group of themes that described the motivations of the students. The importance of motivation in L2 reading has long been recognized in EAP contexts (Komiyama, 2013:150), with Ro (2013:215) stating it ‘plays a vital role in successful L2 reading ability’, and with Asraf & Ahmad (2003:99) considering it to be a key factor in the success of any ER programme, exemplified by an ER study quoted by Robb & Kano (2013:237) that only showed significant proficiency gains with what it considered to be its motivated groups. Motivation can be intrinsic (i.e. the desire to engage in an activity solely for enjoyment, as described by Komiyama, 2013:150) or extrinsic (i.e. the desire to engage in an activity for external rewards, internal feelings and societal values, as also described by Komiyama, *ibid.*), and in the case of the latter, there can be an integrative orientation that is associated with a positive disposition towards the L2 group and the desire to be a part of it (Dornyei & Csizer, 1998:204) and/or an instrumental orientation that is related to the pragmatic benefits of L2 proficiency, such as getting a better job or a higher salary (*ibid.*), with Komiyama (2013:152) quoting studies that offered more detailed constructs for motivation in L2 reading.

Furthermore, De Burgh-Hirabe & Feryok (2013:89) and Kirchhoff (2013:194) reflected that motivation is not static, but instead changes according to how these factors change. Hence with ER, learners continuously need to balance their reading skills with the challenges presented by the reading, otherwise they could end up being bored when their reading skill is stronger, or anxious when the challenges presented by the reading are stronger (Kirchhoff, 2013:197-199), although it would seem that this applies more when they are only intrinsically motivated.

Hence in this study, the results of the In-Depth Interviews (Section-4.7.1 p.177 & Figure-11) showed that the students were highly extrinsically motivated in their studies, which could also shed light on why only slim vocabulary gains (Section-4.2.7 p.147) were recorded during this study’s ER Programme (Section-3.2 p.54) that did not

contribute to their final grade. Hence the students expressed a ‘Strong Motivation for High Grades’, which had been instilled in them at school and in the home. They reported that ‘High Grades are Needed for Good Jobs’, which they felt was necessary to make their families happy, and to give them respect amongst their peers. In addition, they held that ‘Good Jobs Should Have High Salaries’ to support the expenses they considered important, like spending on their parents, their families, their own imminent weddings, new cars, and business ventures.

This profile is not unique to this context, but it has been observed before that students lose enthusiasm for reading when there is no ‘tangible reward of extra hours of credit toward graduation’ (Fitz Gerald, 1947:359), and it has been observed in other countries too, like those in the studies reported by Komiya (2013:151). Hence it would seem that if ER was connected to the students’ final grade, motivation could improve and language acquisition could possibly be enhanced. Robb & Kano (2013:245) went further to insist that there must be ‘an effective way to hold students accountable for their reading’ to successfully implement ER beyond the classroom.

However, Littlejohn (1985:258) noted that, when the participants of his study were asked to use a self-study centre as a course requirement, they quickly discovered that ‘the most important criterion was bodily presence... rather than actual work’ and that they ‘became expert in pretending to do work’ (*ibid.*). Such a situation could easily arise in the context of the students of this study if they were given official marks for their reading during the SSRs, and the situation could be worse if marks were given for the unsupervised reading outside the class. Hence a huge challenge exists in finding an effective way to link ER to the final grade, in order to promote student motivation.

In summary, the above discussion of the results of the In-Depth Interviews (see Section-4.7.1 p.177) yielded many possible factors that could shed light on the modest vocabulary gains (Section-4.2.7 p.147) found after the ER course (Section-3.2 p.54). The most widely reported factor was the students’ lack of ER culture, which included their view that the ER programme was boring, which was different to the view expressed by the participants in most of the surveyed studies. They also included the

conflict they experienced when faced with the ER learning style that differed greatly from their previous social and educational experiences, and their busy academic and social lives that gave little time for reading for pleasure. The participants also reported their strong desire for high academic grades, which they saw as essential to obtain the well-paid jobs they wanted.

Based on all of this, many proposals were discussed to modify future ER courses. These included explicit learning of vocabulary through memorisation, textbooks that actively promote ER, books that are specially chosen by the teacher for the students to read, more in-class silent reading sessions and using media that the students already enjoy (such as the internet and videos) to encourage more reading. Finally, far more fundamental proposals were also discussed in order to improve motivation, such as making the ER course a formal institution-wide part of the curriculum, and such as assigning marks to the reading that would contribute to the students' final grade.

5.2.4 Summary: Response to Research Question 2

Hence to summarise the above sections, it can be suggested in response to RQ2 that the students of this study had the following characteristics that could have been associated with the unspectacular vocabulary gains (see Section-4.2.7 p.147 & Appendix-18) found after the ER programme (described in Section-3.2 p.54):

- In terms of the nature of their reading, they read little in English and even less in Arabic, raising the argument that an Arabic ER programme may also be needed. In addition, the vast majority of their reading was text-book/college related, possibly indicating that they were mainly used to reading when forced to do so, and leading to a discussion on how textbooks could be used to actually encourage ER.

- In terms of the English they had learned outside the college, little of this was learned from English speaking teachers, little was used outside their English lessons, little was used during their travels, they had not prepared themselves for any external English exam, and few had used English at work. All of this indicated their initially low level of

proficiency, informing deliberations into how this affected the low vocabulary gains found.

- In terms of their social and cultural background, they lacked a reading culture, finding reading boring and difficult because it needed skills that they were not used to using. They also had many academic and social activities that left little time for pleasure reading. In addition, they had a strong desire for high academic grades, which they saw as essential to obtain the well-paid jobs they wanted, but they found no such grades were associated with the ER programme, thus possibly affecting their motivation. This rich profile of the students raised many points of discussion relating to how the ER programme could be modified in such a context.

5.3 DISCUSSION OF RESULTS THAT WENT BEYOND THE RESEARCH QUESTIONS

The In-Depth Interviews (Section-4.7.1 Figure-12 p.187) also gave rise to themes that went beyond RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8 p.36), which was a natural result of this part of the study that was more inductive in nature.

Firstly, the 'Presence of Some Limited ER Culture' was reported. The students described that some of their parents read conventional books; that they themselves read via mobile phones; that they enjoyed watching subtitled films; and that they recognized that reading is beneficial for exercising the mind, assisting in their studies and improving their English. This highlights the distinction made by Ro (2013:216) that the 'difference between attitude and motivation is a crucial' because in this study the students had a positive attitude to reading in the sense that did recognize it was useful, but that still did not motivate them to read.

In addition (Section-4.7.1 Figure-12) the 'Presence of Some Limited Intrinsic Motivation' was recorded during the interviews. For example, the students reported that in some instances the grades were not the most important thing to them; that sometimes simply pleasing the teacher was more important; that sometimes they would voluntarily change the settings of their phone applications to English only to improve their English; and that sometimes they would actively use the films they watched to learn new vocabulary through translation.

The fact that the students would translate words from films on their own accord, without being asked to, could be very significant. Nation (2003:4) asserted that learning words through L1 translations is 'most effective', by referring to many studies that supported this, and by going further to state that the criticisms against using L1 translations are not supported by research. He also added that presenting the meaning of a new word through an L1 translation may not be exact, may be indirect, and may take up time that could be better spent using L2, but all of these objections can apply to other ways of

presenting the meaning, such as using pictures, drawings, demonstrations, and realia (Nation, 1978:172). He further referred to many studies that seemed to indicate that using electronic bilingual dictionaries during ER ‘greatly increases the chances of vocabulary learning’ (Nation, 2015:137), because the very short moment spent using the internet or the electronic dictionary does not affect the flow of reading, unlike when using a conventional dictionary. The use of bilingual dictionaries could also help promote L1 and hence enhance the learners’ esteem. Promoting L1 has been explicitly recommended by McCabe (2005:1), while Nation (2003:7) further recommended ‘to avoid doing things that make the L1 seem inferior to English’, which can be affective to the learners. With this background, there could be scope to improve ER with the students of this context, by actively encouraging the use of electronic bilingual dictionaries.

As a result of all the above themes, the participants of the In-Depth Interviews gave many explicit ‘Proposals for Future ER Programmes’ (Section-4.7.1 Figure-12 p.187), given that they expressed that the ER simply needs to be administered in a more attractive way. Taking the ideas of the students makes the research ‘data driven... 'bottom-up', or 'heuristic', generating hypotheses rather than testing them’ (McDonough & McDonough, 1997:79), and any ensuing ideas that are used could be seen as ‘a truly learner-centred approach’ (Littlejohn, 1985:253) because their ideas are used to mould the overall ER programme and the individual activities contained therein.

The students’ call to make ER more interesting directly relates to the factors given by Asraf & Ahmad (2003:99) for a successful ER programme, namely how ‘we are able to motivate the students to read; how the program is organized and implemented; and how their reading is monitored’, and this would give weight to Yamashita’s (2015:174) description of ER as being more of an ‘approach’ that consists of underlying teaching principles that can be applied in many different ways, rather than a ‘method’ that consists of fixed sets of procedures and techniques that must be followed, and it would also reflect the call of Macalister (2008:254) to be flexible in approach when implementing an ER programme.

Many of the participants' explicit 'Proposals for Future ER Programmes' (Section-4.7.1 Figure-12 p.187) were clearly linked to previous themes. For example, they clearly stated that the reading should be linked to their final grade, echoing the previous theme that the students had a Strong Motivation for High Grades (Section-5.2.3 p.206 above and Section-4.7.1 Figure-11 p.184). This is not a strange proposal, even though it goes against the principles of pure ER, given that in many parts of the world students are 'often eager to read during the initial phase of an ER program' (Kirchhoff, 2013:192), but then the realization that it does not explicitly count towards their final grades 'dampens their enthusiasm completely' (Fitz Gerald, 1947:359). This can be applicable to all L2 learners in general, given that L1 learners have no other language and so have an urgent need to learn, while L2 learners already know their own language and so, naturally 'there is a greater dependence on extrinsic (or instrumental) motivation to learn' the new L2 (Elley & Mangubhai, 1983:54).

ER in particular has been noted for being especially challenging for lower level learners (Section-2.5 p.13), such as those in this study, which may mean a low expectation of success and thus a low motivation to read (De Burgh-Hirabe & Feryok, 2013:74), thus potentially making it especially effective in this context to have the reading count towards the final grade. In addition, it has been asserted that using an effective way to hold students accountable for their reading and then administering a final assessment, is essential for the success of ER outside the class (Robb & Kano, 2013:245), but the challenge lies in developing this so that the benefits of ER are not lost.

Participants also explicitly recommended that vocabulary learning activities should accompany the ER (Section-4.7.1 Figure-12 p.187), which echoes the recommendation that ER should be combined with other activities (Section-3.2.3 p.61) like deliberate vocabulary learning (Nation, 2015:142), for which there are many possible activities. Words can be pre-taught, with it being recommended to teach the more frequently occurring words (Nation, 2004:21-22), and this deductive approach could favour more analytic learners (Fortune, 1998:68). Some vocabulary activities need not be in context (Nation, 1995-6:13), while others could be based on the vocabulary found in the books the students are reading during their ER, which leads to a more inductive form of

learning that would favour holistic learners (Fortune, 1998:68). The use of flashcards, which was delightfully described by Elliot (1962:13) over half a century ago, and the use of other methods to aid memorizing vocabulary could be very effective for students who are used to memorizing (Schmitt, 1997:4), such as those in this study who may gain a sense of progress and confidence through memorization, and so may be more motivated, as observed by Sivell (1980:52) with similar students. Finally, vocabulary work can be done through the active skills of speaking (Nation, 1995-6:9-10) and writing, which can provide opportunities to become more fluent (Waring & Nation, 2004:107) when using already-known words.

The students' attraction to vocabulary learning activities could also suggest the usefulness of explicit training in lexical inferencing, which has been defined as the process in which the reader infers the meaning of an unknown word by using any background general knowledge and by considering the available linguistic and contextual clues (Nassaji, 2006:388; Zhang & Koda, 2012:1198; Pulido, 2007:66). It has been argued that inferencing can account for much of the vocabulary learned by both L1 and L2 readers (Nassaji, 2006:388) and that it can enrich the knowledge of partially known words (Nation, 2015:138). However, it has been noted that weaker learners are not so successful when implementing inferencing strategies (Nassaji, 2006:387; Pulido, 2007:68), leading to a call for its explicit training for such students in general (Nassaji, 2006:398) and for Gulf Arab students in particular (O'Sullivan, 2004:6), making this proposal particularly relevant to the context of this study.

Other activities were also suggested (Section-4.7.1 Figure-12 p.187) that use the ER as a 'springboard' for further work (Macalister, 2015:123). Hence some interviewees suggested having discussions on what they had read, echoing the recommendations of Nation (2001:3) and Reid-Thomas & Hill (1993:265). Others proposed summarizing the stories they had read to the rest of the class. Waring (2007:38) observed that stories 'are a part of our human nature' being naturally appealing even to low level learners because they are already accustomed to them. It has though been advised to allow lower level students to prepare first in L1 before doing L2 tasks (Nation, 2003:3), with Asraf & Ahmad (2003:90-93) reporting that when they allowed their students to do so before

retelling to the class the stories they had read, it made the task easier and hence made them more enthusiastic.

Another suggested activity students gave that the ER could lead on to (Section-4.7.1 Figure-12), was reading aloud their written paragraphs, with Day & Bamford (2002:138) also suggesting students reading out in front of the class sections of the stories they had read. Reading aloud (RA) has generally been unpopular in the literature (Gibson, 2008:29) and it has been denounced as ‘a waste of time’ (King, 1978:42) and nothing more than an ‘easy time filler’ (Gibson, 2008:29). RA has been criticized for being boring to the listener (Rounds, 1992:788); for compounding errors in pronunciation if done without teacher correction (Saville-Troike, 1973:400-401); for being difficult and so demotivating (Gibson, 2008:30); and for being unneeded by most learners (*ibid.*). RA has also been censured for slowing down reading and making comprehension more difficult (Griffin, 1992:784), to the extent that it cannot be used at all to check reading comprehension (Rounds, 1992:790; Saville-Troike, 1973:397).

However, it has been argued that RA is simply different to silent reading, and so it should be judged in a different light, given that it involves different skills and offers different advantages (Griffin, 1992:784; Moody, 1974:315). Hence RA has been considered instinctive and reassuring to beginners (Cartledge, 1952:94) and hence necessary for them (Elliot, 1962:14) to build confidence, which is particularly relevant to the context of the learners in this study. It is easier than unaided speech and so it can be used to prepare for conversations (Cartledge, 1952:95) and to rehearse for other speech activities (Gibson, 2008: 31). It has been argued that RA could actually enhance comprehension, because the reader cannot just skim or scan, and instead must consider every word (Moody, 1974:320), and that it may even aid writing (Gibson, 2008:32). In addition, RA has been declared the only practical way to check correct pronunciation when reading (*ibid.*, pp.31-33). Furthermore, RA done by the teacher has been actively encouraged for providing good models of pronunciation and intonation (Cartledge, 1952:96; King, 1978:43; Moody, 1974:317), and it has been found to improve comprehension that may in turn improve attitudes to reading (Amer, 1997:46).

With this background, RA is widely used (Gibson, (2008:29) and even ‘held in high esteem’ (Rounds, 1992: 788) in some parts of the world, and it remains popular with many teachers (Griffin, 1992:785) and students (Gibson, (2008:35), as expressed also by some of the participants in this study. Hence in this context, RA could have a useful role in future ER programmes if used ‘sensitively and appropriately, with clear learning objectives’ (*ibid.*, p.35).

Some of the Proposals for Future ER Programmes related to the Sustained Silent Reading (SSR) sessions. Some participants explicitly said (Section-4.7.1 Figure-12 p.187) that reading the internet or watching videos with English subtitles would make the SSR sessions more interesting, and this was echoed by Meara (2005:5) in the case of watching subtitled videos. However beyond its enjoyment, reading and listening has been considered essential for learning vocabulary (Nation, 1995-6:9), with some noting that incidental learning of L1 vocabulary is mainly learned through these two ways (Pellicer-Sanchez & Schmitt, 2010:33). Also, in the context of the learners in this study, if they find these enjoyable, they are much more likely to continue this kind of ER outside the classroom, which could be very effective in enhancing the benefits of the ER programme.

Another important request made by some participants (Section-4.7.1 Figure-12 p.187) about the SSR sessions was that the teacher should guide them to the kind of books they should read. Once again, this would appear to contradict the important ER characteristic of leaving the students free to choose whatever they want read, running the risk of being boring or difficult to some, as found by De Burgh-Hirabe & Feryok (2013:89) when they chose particular GRs for their ER programme. Hence this would be a modified kind of ER that Day (2015:297) termed ‘*supervised* (or instructed) ER’, with the SSR becoming what Stoller (2015:155) called ‘scaffolded’ SSR. As described before in Section-3.2.2, the teacher can rely on experience to aid the selection of suitable books, and previous studies can also help in this. For example Hafiz & Tudor (1989:10) found that story books were the most popular during their ER programme, and Reid-Thomas & Hill (1993:252) recommended when choosing books that basic matters should not be overlooked, like the cover design, illustrations, page size, font size, and cultural &

geographic setting (see Section-3.2.2 p.57). Macalister (2015:124) also recommended that teachers test the vocabulary knowledge of their students before choosing their books.

Finally, a few Proposals for Future ER Programmes (Section-4.7.1 Figure-12 p.187) related to the use of technology to make filling in the Reading Diary Sheets quicker, less laborious and more accurate, with one participant recommending a particular application that records everything read on the smart phone throughout the whole day. Logs and diaries are perhaps the only way to collect data about activities that span months (Hyland, 2002:188), and in this study, the Reading Diary Sheets (Section-3.6 p.103) were essential to provide details of the participants' reading. However, it was acknowledged that filling them in was time-consuming, and that the recorded information would likely be an underestimate simply because no participant could be expected to remember everything he had read. Hence any electronic method would make the process easier and the information more accurate, and given that mobile phones are a major source of the reading done by young people, using an application to monitor that alone could yield exceptionally rich and useful data.

5.3.1 Summary

In summary, the above discussion concerned the many themes that arose from the In-Depth Interviews (see Section-4.7.1 Figure-12 p.187) that went beyond RQ1 (concerning the impact of the ER programme on vocabulary acquisition) & RQ2 (concerning the impact of other factors on vocabulary acquisition) (see Section-2.8).

Hence, the students described some limited amount of ER, with for example, a few of their family members reading for pleasure, they themselves reading via their phones, they watching subtitled films, and they recognising that reading was a useful skill. It was thus discussed how although the students were unmotivated to read, they did still have a positive attitude to reading, which could be capitalised upon to improve their motivation if, for example, mobile phones and subtitled films (which they already enjoyed using) could be used to encourage more reading. They also described some

occasional instances when they tried to improve their English on their own initiative by, for example, translating new words, which lead to a discussion about the suitability of using L1 translations and dictionaries that contained them.

Finally the students gave many explicit suggestions for future ER programmes, which as discussed, served to be an excellent example of data driven proposals that could be immensely important given that they came directly from arguably the biggest stakeholder: the students. Hence, they put forward ideas such as linking the reading to their final college grades, adding many different activities to the reading, allowing the internet and subtitled videos to be read during the in-class reading sessions, and using technology to simplify and speed up the process of recording the reading diaries, the implications of which were all weighed up, as detailed above in Section-5.3 p.214.

CHAPTER-6

CONCLUSIONS

In the previous chapter, a detailed discussion of the Results was given. In this final chapter, the Conclusions of the project are presented in terms of its contribution to the knowledge, and in terms of its implications for research, practice and their agents.

6.1 CONTRIBUTION TO KNOWLEDGE

This study investigated two research questions, as given before in Section-2.8 p.36, namely:

Research Question 1 What is the impact of an extensive reading (ER)
(RQ1) programme of a relatively short duration, on the
 vocabulary acquisition of male Saudi university students?

Research Question 2 What characteristics of the students, in terms of the nature
(RQ2) of their reading, their other English learning, and their
 culture, may also be associated with any observed
 vocabulary acquisition?

In doing so, it possessed many distinctive characteristics to help address the limitations of earlier surveyed studies that were described in Section-2.7 p.19, namely:

1. A three-month ER programme.
2. An ER programme that allowed a free choice of reading material.
3. A natural learning and teaching context.
4. An ER programme that was incorporated into a normal unaltered course of study.
5. A study of partial vocabulary acquisition.
6. A study of other factors that may have been associated with vocabulary acquisition.
7. A study of poorly motivated, lower-level, young adult male Saudi university students.

Chapter-5 (Discussion) Section-5.1.1 p.199 summarised that in response to RQ1, the ER programme for the above situation had only a small impact on vocabulary acquisition, and in response to RQ2, Section-5.2.4 p.212 summarised many factors that could have contributed to this. In addition, Section-5.3.1 p.220 discussed findings that went beyond both RQ1 and RQ2.

What follows now is a detailed exposition of the valuable contribution that this study can offer to the knowledge of ER, vocabulary acquisition and language acquisition, culminating in a summary in Section-6.1.1 p.230.

Firstly, the results given in Section-4.2.7 Figure-6 p.138 & Appendix-18 showed that there were few statistically significant vocabulary gains after the ER programme (13.1 weeks – see Section-3.2.5 p.66), and these were in the more frequent word levels. In addition, there were greater statistically significant vocabulary gains for the Experimental cohort than the Control, but the difference was not large in terms of absolute numbers of words, and in a few cases the Control actually showed slightly more vocabulary gain.

These unspectacular results could deepen the understanding held about ER, because they seem to challenge those who unconditionally endorse ER as effective and even essential for learning vocabulary, such as Bamford (1984:223) and Nation (1995-6:9) (refer to Section-2.6 p.16). As detailed in the Discussion Section-5.1 p.195, Al-Homoud & Schmitt (2009:396) reported that with their participants, who had a similar social, ethnic and academic background to those in this study, the ER and IR groups showed similarly slight vocabulary gains. This perhaps could lead us to learn that ER is not so effective with the kind of student that was common to both of these studies, qualifying the knowledge presented from the seemingly positive results given in other studies on ER and vocabulary acquisition that used very different kinds of students (see Section-2.7.6 p.34).

Also, in terms of the different aspects of vocabulary knowledge gained, Section-4.2.7 Figure-6 p.138 & Appendix-18 showed that the gains for Spelling (tested through

multiple choice questions) were less than those for Meaning (tested through open translation). Section-5.1 p.195 of the Discussion detailed that this was in contrast to the results of the surveyed studies that showed the results for Spelling recognition were better than those for Meaning through open response (e.g. Pellicer-Sanchez & Schmitt, 2010:41; Pigada & Schmitt, 2006:8; Waring & Takaki, 2003:141), and it was suggested that because the students in this study were Arabic speakers, they found English spellings much harder because Arabic spellings conform almost entirely to their spellings (in contrast to many English spellings, as described by Moody, 1974:318), the Arabic script is completely different to that of English, and Arabic is written in the opposite direction.

This result (that Meaning production was easier than Spelling recognition – see the above paragraph) could contribute to a reassessment of what we know about how vocabulary is learned. Section-3.4.1 p.73 detailed different kinds of vocabulary knowledge, and it quoted the assertion that receptive knowledge is gained before productive knowledge (Elliot, 1962:14; Saville-Troike, 1973:400; Schmitt, 1998:285). However, this study would seem to indicate that the notion that Spelling recognition is easier than Meaning production is misplaced in this setting. Once again, this study's participants were very different to those in the other studies, which could again help to explain this stark contrast.

In addition, as expounded in Section-5.1, the Academic-word section of Test-4 (Meaning through multiple choice questions - detailed in Section-3.4.5 p.86), was expected to be problematic for non-Romance language speakers (Schmitt et al., 2001:69), and to some extent, this was found to be the case in this study, with less vocabulary gain being seen in this section than in the 2000-word and 3000-word sections of the same Test (Section-4.2.7 Figure-6 p.138 & Appendix-18). However, at the same time there was still a greater improvement in Certainty in the Academic-word list than the other two sections, and it was presented in Section-5.1 p.195 that this could have been because the students were being bombarded with a large volume of new academic vocabulary during their first course, and so they were slowly becoming more familiar with them, but could not master their spellings and meanings by the time of the

posttests. Once again, this result from the study would seem to qualify the aforementioned perceived knowledge that Arabic speakers will find it harder to acquire Academic English vocabulary, even though such a notion is entirely plausible given that many academic words are based on languages (e.g. Latin and Greek) that are very different to Arabic.

The recurrent theme of knowledge that emerges from all of this is the importance of the context in which ER is applied. The students in this study were very different to those of many previously surveyed studies – in terms of L1, motivation, culture, reading habits and academic ability – and the one semester was a natural teaching context that was again different to the contexts of many other studies (see Section-2.7.6 p.34). These differences in context could have been critical to the results of this study, and they could be critical in reassessing our knowledge of ER, in implementing ER, and in trying to enhance vocabulary acquisition. As first mentioned in Section-2.7.1 p.20 during the Literature Review, Bax (2003:281-282) observed that methodology is often the dominant strand in teaching training courses, with little emphasis on the myriad of different contexts throughout the world in which those methodologies are to be applied. As a result, he called for context to take at least as important a role as methodology (*ibid.*), and this study would seem to support this call. De Burgh-Hirabe & Feryok (2013:73) similarly recommended that language learners need to be understood as acting within their own contexts, with Ellis (in Kluge, 1997:3) dismissing the notion that experimental research can be generalized to all teaching contexts. Hence, although ER may have been considered successful in a number of contexts around the world, this was not found to be the case in this context, at least in terms of its influence on vocabulary acquisition, significantly qualifying the knowledge presented in the literature about ER and vocabulary acquisition.

Given the importance of context in how learning and teaching is understood, it may be concluded that great pains should always be taken to adapt ER according to the situation in which it is placed. Day (2015) surveyed dozens of ER studies to investigate if their programmes used the 10 teaching principles that he co-recommended over a decade before (Day & Bamford, 2002), and which were largely used when implementing the

ER programme of this study (see Section-3.2 p.54). He termed programmes that used all 10 as ‘pure’ ER, but he found a huge variety of implementation (*ibid.*, p.295). This would lend weight to Yamashita’s suggestion (2015:174) that ER is in practice more of an approach that does have underlying teaching principles, but can be applied in many different ways, as opposed to a rigid method with fixed sets of procedures and techniques that must be followed. Macalister (2008:254) added that there ‘must also be flexibility in approach’ when implementing ER, and in this lies the conclusion that if ER is to be more successful in encouraging vocabulary acquisition in challenging situations like the context of this study, some aspects of pure ER may not be appropriate.

Hence this study’s findings would seem to support a major reconsideration of the knowledge that is associated with ER, in that some aspects of ‘pure’ ER may not be suitable depending on the context, which challenges the bold statements that ‘people learn most of their vocabulary from reading’ (Waring & Takaki, 2003:130) and that ER is ‘the single most effective way’ to improve vocabulary (Bamford, 1984:223), both of which were first quoted in the Literature Review Section-2.6 p.16. Instead, this study’s findings would seem to support the general recommendation of Waring & Nation (2004:106-107) to incorporate direct learning and teaching of vocabulary as being perhaps more applicable to the students of this context, even though such explicit vocabulary learning and teaching are foreign to ‘pure’ ER.

This study further modifies the knowledge of ER and vocabulary acquisition based on what it found with respect to motivation. Asraf & Ahmad (2003:99) viewed motivating students to read as an important part of any ER programme, but given the results of the In-Depth Interviews (Section-4.7.1 p.177) where the students were found to be lacking an ER culture, it would seem that this is even more important in this context. In addition, teacher enthusiasm alone was found to be insufficient to motivate the students (see Section-5.1), and instead the audacious measure of assigning grade-bearing marks to the ER may be needed, as the students themselves suggested during the In-Depth Interviews (Section-4.7.1). Holding students to account for their reading has indeed been viewed a condition for successful unsupervised ER (Robb & Kano, 2013:245), but

the idea of assigning marks to reading for pleasure would seem to be the complete antithesis of the fundamental principle of ER that it should be done for pleasure and for its own sake (Section-2.1 p.5). Hence this study's finding that assigning marks to ER may through enhancing motivation also enhance vocabulary acquisition, requires a fundamental change in how ER is defined, let alone how it is applied.

Another important contribution of this study to knowledge is that a data-driven and inductive approach is extremely important when investigating ER and vocabulary acquisition. Hence in this study, it was only through the Reading Diaries (Section-4.4 p.151 & Section-5.2.1 p.201), the Survey of Exposure to English Learning (Section-4.5 p.156 & Section-5.2.2 p.204), the Supplementary Questioning (Section-4.6 p.162), and the In-Depth Interviews (Section-4.7 p.164 & Section-5.2.3 p.206), that possible reasons for the modest vocabulary gains could be found, addressing RQ2. However, few of the surveyed studies on ER and vocabulary acquisition used such data collection instruments, with the vast majority instead relying on more deductive and experimental methodologies (see Section-2.7.3 p.30) that could not offer as rich and detailed a presentation of what could have been associated alongside ER with any observed vocabulary acquisition.

The data collection instruments described in the above paragraph yielded a further significant contribution to knowledge by giving a rich profile of the students of this context with a level of detail that surpassed that given in any other surveyed study, and as such could be of unique value to other researchers and front-line practitioners dealing with this context. Some important aspects of this profile were (see Section-4.4 p.151, Section-4.5 p.156, Section-4.6 p.162 & Section-4.7 p.164):

- they were mainly in their early twenties;
- they had not typically been taught English in English with native English speaking teachers during any of their school years;
- they read little in L2, and this was almost entirely related to their academic studies;
- they read even less in L1, mainly the internet, subtitled movies and mobile phones;
- few had taken external exams, and those who had, obtained scores that were the equivalent to C-D in the Saudi Arabian educational system;

- few had travelled abroad, and those who had, used little English;
- few had work experience in which they used English;
- they reported that they were not used to reading, they did not like it, they found it burdensome, and they considered it an academic task;
- they reported that they were used to textbook-based learning, memorization of facts, teacher-chosen learning materials, marked assignments, and so they were surprised that the ER Programme did not have any of those features;
- they reported that they had too many lessons to have time to read for pleasure, they missed their family and so they could not focus on their studies well, and they found English difficult to learn, let alone to study other subjects in;
- they reported that many pastimes also took precedence over reading, such as watching films, watching soap operas, watching documentaries, relaxing with friends, & hunting;
- they expressed a strong motivation to get high grades, which had been instilled in them at school and in the home, and that they needed these to get highly paid jobs to make their families happy, to give them respect amongst their peers, and to support the expenses they considered important, like spending on their parents, their families, their own imminent weddings, new cars, and business ventures;
- they did describe a limited amount of reading done by some parents and relatives, and they did recognize that reading is useful for their studies and their English;
- they did describe a few isolated instances of when they were not motivated by grades, but instead by a desire to simply improve or impress;
- and they gave many explicit Proposals for Future ER Programmes (Section-4.7.1 Figure-12 p.187), explaining that the ER simply needs to be presented in a more attractive way.

This detailed description of the participants is another example of where a ‘bottom-up’ or ‘heuristic’ approach (as termed by McDonough & McDonough, 1997) yielded information that was not previously envisaged. In this study, such information went beyond RQ1 & RQ2, such as the reasons why the students were so highly extrinsically motivated, and such as the detailed suggestions the students offered for improving the ER programme (see Section-4.7.1 p.177).

6.1.1 Summary

Overall, this study has offered a valuable contribution to knowledge in the many fundamental and detailed ways that are comprehensively described above. These are now summarised below:

- Challenging the unqualified assertion that ER is effective and even necessary for learning vocabulary.
- Questioning the notion that receptive vocabulary knowledge is always acquired before productive knowledge.
- Purporting that the context in which ER takes place is a critical factor that must be taken into account when investigating and implementing it.
- Proposing that for the context of the students of this study, pure ER may not be suitable, and instead a heavily modified version may be more appropriate, even though some of the modifications described can be considered audacious and even contrary to at least some principles of ER.
- Highlighting the effectiveness of using data-driven inductive research methods to allow unexpected findings to emerge.
- Providing an in-depth profile of the students of this context, the detail of which surpasses that given in any other surveyed study.

All of this added to distinctiveness of this study, when placed amongst the knowledge presented by others in the field.

6.2 IMPLICATIONS FOR RESEARCH & PRACTICE

Many recommendations can be made based on the findings of this study that related to both research and practice, and hence by extension to both researchers and practitioners, with the researcher holding the view that all these four are intertwined, mutually inclusive and inseparable, as explained in Chapter-1 (Introduction). Hence in this final section, the implications of this study are presented with all four alongside each other, emphasising that research is limited if cannot offer guidance for practice, and emphasising that both research and practice are impossible without its active agents, closing with a summary of final thoughts in Section-6.2.1 p.237.

Arguably the most fundamental implication emerging from this study is that researchers and teachers alike need to be more aware of context when applying methodologies, as recommended before by Bax (2003), De Burgh-Hirabe & Feryok (2013:73) and others. ER has been recommended by many to enhance proficiency in general (Section-2.4 p.8) and vocabulary acquisition in particular (Section-2.6 p.16), but in the specific context of this study, the young poorly motivated lower-level Saudi students, who read little even in L1, and who were engaged in a single semester university preparatory course, did not seem to show much in the way of significant vocabulary gains during their ER programme (Section-4.2.7 Figure-6 p.138 & Appendix-18). Hence a greater attention is needed, as described by Hu (2005:331), when ‘thinking globally’ based on ideas published in research from around the world and then ‘acting locally’ based on the specific teaching context, and this as he further described, requires understanding and experience of the students in hand (*ibid.*).

Many have pointed that ER, despite its challenges, remains an extremely beneficial activity (Brown, 2009:240) for improving more than just vocabulary acquisition (Waring & Takaki, 2003:155), and in this study the participants did acknowledge it was useful albeit at the same time uninteresting (Section-4.7.1 Figure-12 p.187). Hence, in this context, it may be the case that, using the terminology of Day (2015:296-297), the ER needs to be ‘modified’ i.e. using ‘many’ of the ER principles recommended by Day & Bamford (2002), or even it may need to be ‘light’ i.e. using just ‘some’ of those

principles. The need to adapt seemingly ideal methodologies is not new in the history of language teaching, as pointed out by Klapper (2003:33) who further explained that no amount of research evidence can make a method applicable in all times, in all places and for all learners. It could thus be further recommended, as done so by Yamashita (2015:174) and Macalister (2008:254), that ER should be considered as more of a flexible approach that consists of underlying teaching principles that is then applied in different ways, as opposed to a method of fixed procedures that must be rigidly followed.

With this in mind, this study yields many important implications that can affect how ER should be modified for this specific context of young poorly motivated lower-level Saudi students, who read little even in L1, and who were engaged in just a single semester of a university preparatory course. A key recommendation is that more research is needed to unpick how to motivate students in this challenging context to read more and to enjoy reading more, given that enjoyment is a key feature of ER (Yamashita, 2015:173), and given that although teacher enthusiasm is considered a key factor in successful ER programmes (Asraf & Ahmad, 2003:99), it was not sufficient in this case. Linking the ER to marks that count towards the students' final grade could be a good direction to take, given the highly extrinsic motivation displayed by the participants (Section-4.7.1 Figure-11 p.184), and given that they themselves advocated this in order to improve the ER course (Section-4.7.1 Figure-12 p.187), even though such a measure is not a characteristic of pure ER.

However, with this recommendation comes the challenging call for a deeper layer of research into how to effectively link grades to ER, i.e. how to assign grades to ER and yet still retain its important features. Certain aspects of ER, such as the number of books read in class and the speed of such reading could be assigned a summative grade, but as highlighted previously in Section-5.2 p.201, students can be 'expert at pretending to do work' (Littlejohn, 1985:258) and so recording the quantity and speed of reading may have no relevance whatsoever to intangible features like concentration, interest and enjoyment that are thought to drive ER to enhance language proficiency and vocabulary acquisition.

Further to this, it can be recommended that ER is made a formal integral part of the course. This has been already advocated by many (see Section-2.4 p.8), but in this context, the need is greater. The participants of this study expressed strong extrinsic motivation (Section-4.7.1 & Figure-11 p.184) and so upgrading ER from an optional extra activity of some teachers, to a formal obligatory task that is implemented throughout all classes, could raise ER's profile and importance in the eyes of the students, and hence make them approach it more seriously and positively. Coupled with this is the recommendation to use text books that actively encourage ER through dedicated chapters or through a dedicated part in every chapter. The students reported that they had been used to text-book based learning throughout their entire academic life (Section-4.7.1 Figure-10 p.179), and so it could be likely that in this context, ER's explicit inclusion in their text-book would, as suggested by Brown (2009:240) give ER more 'credibility and legitimacy', and hence attract better attitudes towards it.

The implications of this study also extend to the individual activities that can be added to future ER programmes, which may be beneficial for this particular context, even though they are not a feature of pure ER. One such kind of activity is explicit vocabulary teaching, given the modest vocabulary gains found (Section-4.2.7 Figure-6 p.138 & Appendix-18) during this study's ER programme that had no explicit vocabulary teaching (Section-3.2 p.54). The students themselves also both voiced a liking for vocabulary activities and voiced a request that these should be made a part of the ER (Section-4.7.1 Figure-12 p.187), which would fit with the recommendation of Nation (2015:142) to combine ER with deliberate vocabulary learning. As an extension of this, explicit training in lexical inferencing strategies may also be useful for students in the context, as detailed earlier in Section-5.3 p.214, and as recommended by O'Sullivan (2004:6) for other Gulf Arab students.

Other activities that can be recommended for future ER programmes are discussions, presentations and read-aloud summaries, which were suggested by the students themselves (Section-4.7.1 Figure-12 p.187). Reading Aloud (RA) could be particularly useful in this context, as discussed earlier in Section-5.3 p.214, given the lower-level

ability of the students. Hence RA could be used as a preparation stage before the more challenging presentations and discussions (Cartledge, 1952:95; Gibson, 2008: 31), and RA could provide an excellent opportunity for the teacher to assist the students in their pronunciation (Gibson, 2008:31-33). Furthermore, it could also be used to assess the ER, by allocating marks to the quality of the students' RA as they progress during the weeks of the programme, which in turn could increase motivation in the extrinsically motivated students of this context.

Recommendations can also be made for the reading materials used during the ER programme. In this study, an argument for giving a completely free choice of books was presented, and this was adopted during the ER programme (Section-3.2.2 p.57). However, given the lack of large vocabulary gains during the ER programme (Section-4.2.7 Figure-6 p.138 & Appendix-18) and given the explicit proposals of the students themselves (Section-4.7.1 Figure-12 p.187), it could be recommended that in this context the teacher would need to take a greater role in guiding the students to appropriate books. This once again is not a feature of pure ER, but instead could be described as '*supervised* (or instructed) ER' (Day, 2015:297), that could be necessary for more successful ER in this context. It would be advisable in such a situation to test the students on their initial vocabulary knowledge to determine the appropriate level of difficulty of the books, as suggested by Macalister (2015:124), and there is no obvious reason why graded readers (GRs) should not be given as least as much encouragement as unsimplified material, while some would argue that it is essential for GRs to be encouraged (Nation, 2015:140).

Another key implication, based on the modest vocabulary gains during the semester's ER programme, is that the ER programme should be much longer. Kirchhoff (2013:193) recommended ER should be for more than just one semester, and Asraf & Ahmad (2003:99) said it should be for more than a year to yield measurable improvements. However this would in turn require the ambitious call for a wholesale shift in attitudes towards ER amongst managers of language courses, because in most cases (such as in this study) ER is just a short term extra activity that is fitted in by a few individual teachers (see Section-2.5 p.13). Furthermore, this is a challenge that must be taken up

by all proponents of ER – from academics to front-line teachers – if longer term programmes are to be realised.

A number of implications also relate to the in-class sustained silent reading (SSR) sessions. In this study, although every effort was made to hold them every day, in the end only an average of 3.53 were possible (Section-3.2.3 p.61) each week. Hence it should be first recommended that the SSR sessions are held every day, as advocated by Macalister (2015:127) as well, although this would probably require the active support of the management of educational establishments, which could in itself be difficult to achieve. SSR sessions are indeed a time when reading is forced upon the students and as such, conflict with the voluntary reading principle of ER. However, in the context of the lower-level students of this study, and given that they themselves reported many activities that restricted the opportunities to read when outside the classroom (Section-4.7.1 Figure-10 p.179), more SSR sessions would seem beneficial to give them at least some of the dedicated reading time they need to hopefully develop their own voluntary reading habit.

It could also be recommended that the SSR sessions should allow reading of the internet and watching subtitled videos. These were explicitly requested by the students during the In-Depth Interviews (Section-4.7.1 Figure-12 p.187) and they were reported through the Reading Diaries (Section-4.4.1 p.154) to be two of their favourite past times, although in L1. Day (2015:297) further envisaged that the internet would be a major source of reading for students outside the class, and this could be very much the case for the context of these students who enjoy using the internet in L1 and so potentially can use it too for L2 reading. Similarly, the SSR sessions can also encourage reading books accompanied with listening to their audio files, which has been recommended for many years (Williams, 1986:45), making it desirable for more research to be conducted into this for the context of the students in this study.

A bold recommendation that could also aid the effectiveness of the English ER programme is to actually implement an Arabic ER programme either beforehand or alongside it. The In-Depth Interviews (Section-4.7.1 Figure-10 p.179) and the Reading

Diaries (Section-4.4.1 p.154) showed that the students of this context read even less in L1 than they did in L2, and even the L2 reading was modest and almost entirely related to their formal academic studies. Hence the very skill of frequent reading for pleasure was new to them. Yamashita (2004:13-14) reported that L1 reading attitudes do transfer to those for L2, which prompted the recommendation to encourage L1 reading in order to promote L2 reading, which could be particularly relevant to this context. However, such a recommendation needs coordination with the L1 language learning department, which may not even exist in many English language learning institutes, and even if it did, such a proposal would need the backing of the administration, which could only be reasonably secured if a convincing argument can be put forward. Hence another implication that is associated with this and that even precedes this, is that more research into the effectiveness of such an audacious proposal is required.

Broader recommendations can be made relating to future studies on vocabulary acquisition and ER. In this study, the results of the vocabulary tests that involved different levels of certainty (i.e. Correctness Alone, Correctness Factored by Level of Certainty, and Certainty Alone) showed that none of these three showed consistently better statistically significant gains than the others (Section-4.2.6 p.145 & Section-4.2.7 p.147). It is possible that a longer ER programme would lead to more noticeable vocabulary gains which would then allow a greater differentiation between the different levels of certainty to be observed. However, more research in general is still needed for the different aspects of partial vocabulary acquisition that could occur during an ER programme. As done in this project, that research should include different aspects of word knowledge (such as spelling, meaning, etc.) and different levels of certainty. In addition, more research is also needed into the long term retention of both new vocabulary learned and existing vocabulary that is reinforced when engaged in an ER programme, which is a call that has been made before by some such as Waring & Takaki (2003:130-133).

Furthermore, it can be recommended that more open, inductive and exploratory research is needed when studying ER and vocabulary acquisition. This study greatly benefited from this, because it was only through its exploratory data collection tools such as the

Reading Diaries (Section-4.4 p.151 & Section-5.2.1 p.201), the Survey of Exposure to English Learning (Section-4.5 p.156 & Section-5.2.2 p.204), the Supplementary Questioning (Section-4.6), and the In-Depth Interviews (Section-4.7 p.164 & Section-5.2.3 p.206), that possible reasons for the modest vocabulary gains could be found, addressing RQ2. Also, it was only through this exploratory approach that findings were discovered that went beyond the research questions and that were not envisaged beforehand, such as possible reasons for the highly extrinsically motivated nature of the participants, and such as the many suggestions to improve future ER courses that were explicitly voiced by them. In order to achieve this, modern electronic methods can be recommended to record richer and more detailed information, such as the mobile phone application that was described by one of the participants of the In-Depth Interviews (Section-4.7.1 Figure-12 p.187), which could make tasks like filling-in the Reading Diary Sheets quicker, less laborious and more accurate. Given that mobile phones are arguably the main vehicle for pleasure reading done by young people, using an application to monitor this could in itself yield exceptionally rich and valuable data.

6.2.1 Summary & Final Thoughts

In summary, this study raised the following important implications for research, practice and their agents:

- Researchers and practitioners need a greater awareness of context when investigating and applying methodologies.
- More research is needed to investigate how to motivate the students in the challenging context of this study.
- ER programmes in this context can be more effective if they continue for much longer than just one semester, if they are a formal and integral part of the overall English course, and if they are encouraged with text books that explicitly promote ER.
- ER programmes in this context can also be more effective if the in-class silent reading sessions take a more prominent role. This includes ensuring that they are held every day, and it includes actively using them to encourage reading the internet and subtitled videos.

- ER should be viewed as a flexible approach that can be adapted according to the context.
- ER in the context of the students of this study may need important and even radical modifications. These include using teacher-selected books, adding extra activities, assigning grade bearing marks for the reading, and even implementing an additional Arabic ER course.
- More research is required to determine best practice when implementing the above modifications to ER programmes in this context, and more research is needed to examine their effectiveness.
- More studies are needed to investigate other aspects of vocabulary acquisition in this context.
- There is an important need for more research on vocabulary acquisition and ER that uses inductive and exploratory methods, to allow for unexpected findings to emerge.

In terms of final thoughts, researchers cannot remain immersed in their research activities only, but instead they must go beyond by promoting their research to all stakeholders in English language learning. Some have expressed that getting teachers to implement research findings is the major challenge (Waring & Nation, 2004:108; Kluge, 1997:2), however in terms of ER being implemented with the students of this context, convincing teachers about ER is not enough, and instead this study has found that implementing ER depends on the attitudes of all of those involved. This covers the students & their parents who may find the mere concept of reading for pleasure strange and boring, let alone the idea that such a non-academic activity could reap huge benefits in their language learning. And this also covers the managers & owners of language learning institutions who may need a great deal of convincing before they can invest the large amounts of time and resources that are required for the long term and large scale implementation that ER requires.

This is where the real challenge lies: promoting the findings of research to those who may not even know of its very existence.

BIBLIOGRAPHY

Al-Homoud, F. & Schmitt, N. (2009). 'Extensive reading in a challenging environment: a comparison of extensive and intensive reading approaches in Saudi Arabia' *Language Teaching Research* 13, pp.383-401.

Allan, R. (2009). 'Can a graded reader corpus provide 'authentic' input?' *ELT Journal* 63 (1), pp.23-32.

Amer, A.A. (1997). 'The effect of the teacher's reading aloud on the reading comprehension of EFL students' *ELT Journal* 51 (1), pp.43-47.

Arnold, W. (2008). 'Young EFL Pupils Reading Multicultural Children's Fiction: an Ethnographic Case Study in a Swedish Language Primary School in Finland' *ELT Journal* 62 (1), pp.96-100.

Aronson, J. (1995). 'A Pragmatic View of Thematic Analysis' *The Qualitative Report* 2(1), pp.1-3.

Asraf, R.M., & Ahmad, I.S. (2003). 'Promoting English language development and the reading habit among students in rural schools through Guided Extensive Reading program' *Reading in a Foreign Language* 15 (2), pp.83-102.

Azmuddin, R.A., Ali, Z., Ngah, E., Tamili, L.M. & Ruslim, N.M. (2014). 'Extensive Reading using Graded Readers' *International Journal of Research in Social Sciences* 3 (8), pp.109-113.

Bamford, J. (1984). 'Extensive Reading by means of Graded Readers' *Reading in a Foreign Language* 2 (2), pp.218-260.

Bauer, L. & Nation, P. (1993). 'Word Families' *International Journal of Lexicography* 6 (4), pp.253-279.

Bax, S. (2003). 'The end of CLT: a context approach to language teaching' *ELT Journal* 57 (3), pp.278-287.

Beale, J. (2002). 'Is communicative language teaching a thing of the past?' *Babel* 37 (1), pp.12-16. Available from: http://www.jasonbeale.com/essaypages/clt_essay.html [accessed 2011]

Beglar, D. & Hunt, A. (2005). 'Six principles for teaching foreign language vocabulary: A commentary on Laufer, Meara and Nation's 'Ten Best Ideas'' *The Language Teacher* 29 (7), pp.7-10.

Beglar, D., Hunt, A. & Kite, Y. (2012). 'The Effect of Pleasure Reading on Japanese University EFL Learners' Reading Rates' *Language Learning* 62 (3), pp.665-703.

Bell, T. (2001). 'Extensive Reading: Speed and Comprehension' *The Reading Matrix* 1 (1). Available from: <http://www.readingmatrix.com/articles/bell/article.pdf> [accessed 2016].

Benson, M.J. (1991). 'University ESL reading: A content analysis' *English for Specific Purposes* 10 (2), pp.75-88.

Benson, P. (2001). *Teaching and Researching Autonomy in Language Learning*. Harlow: Longman.

Benson, P. (2004). '(Auto)biography and Learner Diversity' in Benson, P. & Nunan, D. (eds.) *Learners' Stories: Difference and Diversity in Language Learning*. Cambridge: Cambridge University Press, pp.4-21.

Bitchener, J., Young, S. & Cameron, D. (2005). 'The effect of different types of corrective feedback on ESL student writing' *Journal of Second Language Writing* 14, pp.191-205.

Bloomer, A., Griffiths, P. & Merrison, A.J. (2005). *Introducing Language in Use: A Course Book*. Oxford: Routledge .

Bradley, E., Curry, L. & Devers, K. (2007). 'Qualitative Data Analysis for Health Services Research: Developing Taxonomy, Themes, and Theory' *Health Services Research* 42 (4), pp.1758-1772.

Braun, V. & Clarke, V. (2006) 'Using thematic analysis in psychology' *Qualitative Research in Psychology* 3 (2). pp.77-101.

Brezina, V. & Gablasova, D. (2013). 'Is There a Core General Vocabulary? Introducing the New General Service List' *Applied Linguistics Advance Access*, pp.1-23. Available from:
<http://applied.oxfordjournals.org/content/early/2013/08/25/applin.amt018.full.pdf+html>
[accessed 2014].

British Association for Applied Linguistics (BAAL) (2010). *Recommendations on Good Practice in Applied Linguistics*. Available from:
http://www.baal.org.uk/about_goodpractice_full.pdf [accessed 2010].

Broughton, G. (1962). 'Don't Shoot the Editor' *ELT Journal* 16 (4), pp.199-204.

Brown, D. (2009). 'Why and how textbooks should encourage extensive reading' *ELT Journal* 63 (3), pp.238-245.

Brown, M. (2013). 'Modern life means children miss out on pleasures of reading a good book' *The Guardian*. Available from:

<https://www.theguardian.com/books/2013/jul/06/children-miss-out-on-reading>
[accessed 2016].

Burns, A. (1999). *Collaborative Action Research for English Language Teachers*.
Cambridge: Cambridge University Press.

Cambridge English Language Assessment (2014). 'Cambridge English: Preliminary (PET)' Available from: <http://www.cambridgeenglish.org/preliminary/>, [accessed 2014]

Carrell, P.L. & Carson, J.G. (1998). 'Extensive and intensive reading in an EAP setting' *English for Specific Purposes* 16 (1), pp.47-60.

Cartledge, H.A. (1952). 'Reading Aloud' *ELT Journal* 6 (3), pp. 94-96.

Chappelle, C.A & Duff, P.A. (2003). 'Some Guidelines for Conducting Quantitative and Qualitative Research in TESOL' *TESOL Quarterly* 37 (1), pp.157-178.

Chun, E., Choi, S., & Kim, J. (2012). 'The effect of extensive reading and paired-associate learning on long-term vocabulary retention: An event-related potential study' *Neuroscience Letters* 521 (2), pp.125-129.

Claridge, G. (2012). 'Graded readers: How the publishers make the grade' *Reading in a Foreign Language* 24 (1), pp.106-119.

Clark, M.K. & Ishida, S. (2005). 'Vocabulary knowledge differences between placed and promoted EAP students' *Journal for English for Academic Purposes* 4, pp.225-238.

Clarke, D.F. & Nation, I.S.P. (1980). 'Guessing the meanings of words from context: Strategy and techniques' *System* 8 (3), pp.211-220.

Cotterall, S. (2000). 'Promoting learner autonomy through the curriculum: principles for designing language courses' *ELT Journal* 54 (2), pp.109-117.

Cutting, J. (2008). *Pragmatics and Discourse: A Resource Book for Students*. 2nd Edition. London: Routledge.

Davies, D. (2005). *Varieties of Modern English*. Harlow: Pearson Longman.

Day, R. (2007). 'Learning to read in a foreign language' *The Language Teacher* 31 (7), pp.19-21.

Day, R. (2015). 'Extending extensive reading' *Reading in a Foreign Language* 27 (2), pp.294–301.

Day, R. & Bamford, J. (2002). 'Top Ten Principles for Teaching Extensive Reading' *Reading in a Foreign Language* 14 (2), pp.136-141.

De Burgh-Hirabe, R. & Feryok, A. (2013). 'A model of motivation for extensive reading in Japanese as a foreign language' *Reading in a Foreign Language* 25 (1), pp.72–93.

Dörnyei, Z. (2003). *Questionnaires in second language research: construction, administration and processing*. New Jersey: L. Erlbaum.

Dörnyei, Z. (2007). *Research Methods in Applied Linguistics*. Oxford: Oxford University Press.

Dörnyei, Z. & Csizer, K. (1998). 'Ten commandments for motivating language learners: results of an empirical study' *Language Teaching Research* 2 (3), pp.203-229.

Dörnyei, Z. & Thurrell, S. (1994). 'Teaching conversational skills intensively: course content and rationale' *ELT Journal* 48 (1), pp.40-49.

Dufficy, P. (2004). 'Predisposition to choose: the language of an information gap task in a multilingual primary classroom' *Language Teaching Research* 8 (3), pp.241-261.

Edinburgh Project on Extensive Reading (EPER) (1992). 'Guide to Organising Programmes of Extensive Reading' *Institute for Applied Language Studies at the University of Edinburgh*. Available from: <http://er-central.com/wp-content/uploads/EPER-Guide-1992.pdf> [accessed 2016].

Elley, W.B. & Mangubhai, F. (1983). 'The Impact of Reading on Second Language Learning' *Reading Research Quarterly* 19 (1), pp.53-67.

Elley, W.B. (2000). 'The Potential of Book Floods for Raising Literacy Levels' *International Review of Education* 46, pp.233–255.

Elliot, A.V.P. (1962). 'The Reading lesson' *ELT Journal* 17 (1), pp.9-16.

Errey, L., Cho, N.H. & Ota, T. (2004). 'Learning from Weblogs: a tale of two students.' *The East Asian Learner* 1 (2), pp.5-12.

Evans, S. & Green, C. (2007). 'Why EAP is necessary: A survey of Hong Kong tertiary students' *Journal of English for Academic Purposes* 6, pp.3-17.

Extensive Reading Foundation (ERF) (2016). *ER Bibliography*. Available from: <http://erfoundation.org/wordpress/er-bibliography/> [accessed 2016].

Farrell, T.S.C. (2001). 'Tailoring Reflection to Individual Needs: a TESOL case study' *Journal of Education for Teaching* 27 (1), pp.23-38.

Fereday, J. & Muir-Cochrane, E. (2006). 'Demonstrating Rigor Using Thematic Analysis: A Hybrid Approach of Inductive and Deductive Coding and Theme Development' *International Journal of Qualitative Methods* 5 (1), pp.80-92.

Ferris, D.R. (1997). "The Influence of Teacher Commentary on Student Revision" *TESOL Quarterly* 31 (2), pp.315-339.

Fitz Gerald, T.A. (1947). 'A Plea for Graded Reading Texts' *The Modern Language Journal* 31 (6), pp.359-362.

Fortune, A. (1998). 'Survey review: grammar practice books' *ELT Journal* 52 (1), pp.67-80.

Freeman, D. & Johnson, K.E. (1998). 'Reconceptualizing the Knowledge-Base of Language Teacher Education' *TESOL Quarterly* 32 (3), pp.397-417.

Gibson, S. (2008). 'Reading aloud: a useful learning tool?' *ELT Journal* 62 (1), pp.29-36.

Gray, J. (2000). 'The ELT coursebook as a cultural artifact: how teachers censor and adapt' *ELT Journal* 54 (3), pp.274-283.

Graves, K. (2000). *Designing Language Courses*. Boston: Heinle & Heinle.

Griffin, S.M. (1992). 'Reading Aloud. An Educator Comments' *TESOL Quarterly* 26 (4), pp.784-787.

Hafiz, F.M. & Tudor, I. (1989). 'Extensive reading and the development of language skills' *ELT Journal* 43 (1), pp.4-13.

Hammond, J. (2006). 'High challenge, high support: Integrating language and content instruction for diverse learners in an English literature classroom' *Journal of English for Academic Purposes* 5, pp.269-283.

Hamphire, S. & Anoro, M.A.A. (2004). 'The siren call of the task' *ELT Journal* 58 (1), pp.71-74.

- Harwood, N. (2005). 'What do we want EAP teaching materials for?' *Journal of English for Academic Purposes* 4 (2), pp.149-161.
- Hewitt-Taylor, J. (2001) 'Use of constant comparative analysis in qualitative research' *Nursing Standard* 15, 42, pp.39-42.
- Hill, D.R. (1997). 'Survey review: Graded readers' *ELT Journal* 51 (1), pp.57-81.
- Hill, D.R. (2008). 'Survey review: Graded readers in English' *ELT Journal* 62 (2), pp.184-204.
- Hill, D.R. & Reid-Thomas, H. (1988). 'Survey review: Graded readers (Part 1)' *ELT Journal* 42 (1), pp.44-52.
- Hirsh, D. & Nation, P. (1992). 'What Vocabulary Size is Needed to Read Unsimplified Texts for Pleasure?' *Reading in a Foreign Language* 8 (2), pp.689-696.
- Holliday, A. (1994). 'The house of TESEP and the communicative approach: the special needs of state English language education' *ELT Journal* 48 (1), pp.3-11.
- Holliday, A. (2002). *Doing and Writing Qualitative Research*. London: Sage.
- Holliday, A. (2005). *The Struggle to Teach English as an International Language*. Oxford: Oxford University Press.
- Holliday, A. (2006). 'Native-speakerism' *ELT Journal* 60 (4), pp.385-387.
- Holliday, A. (2007). 'Response to 'ELT and 'the spirit of the times'' *ELT Journal* 61 (4), pp.360-366.

- Hu, G. (2005). 'Using peer review with Chinese ESL student writers' *Language Teaching Research* 9 (3), pp.321-342.
- Hughes, A. & Lascaratou, C. (1982). 'Competing criteria for error gravity' *ELT Journal* 36 (3), pp.175-182.
- Hyland, K. (2002). *Teaching and Researching Writing*. Harlow: Longman.
- Inzumi, S. (2002). 'Output, input enhancement, and the noticing hypothesis' *Studies in Second Language Acquisition* 24, pp.541-577.
- Ishii, T. & Schmitt, N. (2009). 'Developing an Integrated Diagnostic Test of Vocabulary Size and Depth' *RELC Journal* 40 (5), pp.5-22.
- James, C. (1977). 'Judgements of Error Gravities' *ELT Journal* 31 (2), pp.116-124.
- Janopoulos, M. (1986). 'The Relationship of Pleasure Reading and Second Language Writing Proficiency' *TESOL Quarterly* 20 (4), pp.763-768.
- Joffe, H. (2012). 'Thematic Analysis' in Harper, D. & Thompson, A. (eds.) *Qualitative Research Methods in Mental Health and Psychotherapy: A Guide for Students and Practitioners*. Chichester: Wiley-Blackwell, pp.209-223.
- Kachru, B.B. (1995). 'World Englishes: Approaches, Issues and Resources' in Brown, H.D. & Gonzo, S. (eds.) *Readings in Second Language Acquisition*. Englewood Cliffs, New Jersey: Prentice Hall.
- Karlin, O., & Romanko, R. (2010). 'Examining multiple variables within a single ER setting' *The Reading Matrix* 10, pp.181-204.
- King, A.H. (1978). 'Some General Principles of Advanced Reading Instruction' *ELT Journal* 33 (1), pp.38-45.

- Kirchhoff, C. (2013). 'L2 extensive reading and flow: Clarifying the relationship' *Reading in a Foreign Language* 25 (2), pp.192–212.
- Klapper, J. (2003). 'Taking communication to task? A critical review of recent trends in language teaching' *Language Learning Journal* 27, pp.33-42.
- Kluge, D. (1997). 'Interview with Rod Ellis' *The Language Teacher Online* Issue 21.12 - December 1997. Available from: <http://jalt-publications.org/tlt/files/97/dec/kluge.html>, [accessed 2009].
- Komiyama, R. (2013). 'Factors underlying second language reading motivation of adult EAP students' *Reading in a Foreign Language* 25 (2), pp.149–169.
- Kyongho, H. & Nation, P. (1989). 'Reducing the Vocabulary Load and Encouraging Vocabulary Learning through Reading Newspapers' *Reading in a Foreign Language* 6 (1), pp.323-335.
- Laufer, B. (2005). 'Ten best ideas for teaching vocabulary' *The Language Teacher* 29 (7), pp.3-4.
- Landsberger, H.A. (1958). *Hawthorne Revisited*. New York: Ithaca.
- Lederman, L.C. (1990). 'Assessing Educational Effectiveness: The Focus Group Interview as a Technique for Data Collection' *Communication Education* 38, pp.117-127.
- Li, J. & Schmitt, N. (2009). 'The acquisition of lexical phrases in academic writing: A longitudinal case study' *Journal of Second Language Writing* 18, pp.85-102.
- Littlejohn, A. (1985). 'Learner choice in language study' *ELT Journal* 39 (4), pp. 253-261.

Littlewood, W. (2000). 'Do Asian students really want to listen and obey?' *ELT Journal* 54 (1), pp.31-36.

Littlewood, W. (2001). 'Students' attitudes to classroom English learning: a cross-cultural study' *Language Teaching Research* 5 (1), pp.3-28.

Lott, B. (1960). 'Graded and Restricted Vocabularies and their Use in the Oral Teaching of English as a Second Language - II' *ELT Journal* 14 (2), pp.65-70.

Macalister, J. (2008). 'Implementing extensive reading in an EAP programme' *ELT Journal* 62 (3), pp.248-256.

Macalister, J. (2015). 'Guidelines or commandments? Reconsidering core principles in extensive reading' *Reading in a Foreign Language* 27 (1), pp.122–128.

Martyn, E. (2000). 'Syllabus negotiation in a school of nursing' in Breen, M. & Littlejohn, A. *Classroom Decision Making*. Cambridge: Cambridge University Press, pp.150-162.

McCabe, R.V. (2005). 'Tailoring EAP materials to the English Second Language needs of tertiary students from multilingual indigenous backgrounds' Paper presented at the Conference on Bi-/Multilingual Universities, Helsinki University, Finland, 1-3 September, 2005. Available from: <http://www.palmenia.helsinki.fi/congress/bilingual2005/presentations/McCabe.pdf> [accessed 2011].

McDonough J. & McDonough S.M. (1997). *Research Methods for English Language Teachers*. London: Arnold.

Meara, P. (2005). 'Ten best ideas for teaching vocabulary' *The Language Teacher* 29 (7), pp.4-5.

- Mermelstein, A.D. (2015). 'Improving EFL learners' writing through enhanced extensive reading' *Reading in a Foreign Language* 27 (2), pp.182–198.
- Moody, H.L.B. (1974). 'Technique and Art in Reading Aloud' *ELT Journal* 24 (4), pp. 315-324.
- Nakanishi, T. (2015). 'A Meta-Analysis of Extensive Reading Research' *TESOL Quarterly* 49 (1), pp.6-37.
- Nakatsuhara, F. (2004). 'What May Teachers Need to Take into Consideration, When Teaching The Present Perfect with Coursebooks?' *Essex Graduate Student Papers in Language and Linguistics* 6, pp.55-74. Available from: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.144.5946&rep=rep1&type=pdf> [accessed 2015].
- Nassaji, H. (2006). 'The Relationship Between Depth of Vocabulary Knowledge and L2 Learners' Lexical Inferencing Strategy Use and Success' *The Modern Language Journal* 90 (3), pp.387–401.
- Nation, I.S.P. (1978). 'Translation and the Teaching of Meaning: Some Techniques' *ELT Journal* 32 (3), pp.171-175.
- Nation, I.S.P. (1983). 'Testing and Teaching Vocabulary' *Guidelines* 5 (1), pp.12-25.
- Nation, I.S.P. (1995-6). 'Best practice in vocabulary teaching and learning' *EA Journal* 13 (2), pp.7-15.
- Nation, I.S.P. (2001). 'Planning and running an extensive reading program' *NUCB Journal of Language Culture and Communication* 3 (1), pp.1- 8.

Nation, I.S.P. (2003). 'The role of the first language in foreign language learning' *Asian EFL Journal* 5 (2). Available from: http://www.asian-efl-journal.com/june_2003_pn.pdf [accessed 2009].

Nation, I.S.P. (2004). 'Vocabulary Learning and Intensive Reading' *EA Journal* 21 (2), pp.20-29.

Nation, I.S.P. (2005a). 'Ten best ideas for teaching vocabulary' *The Language Teacher* 29 (7), pp.5-6.

Nation, I.S.P. (2005b). 'Vocabulary Learning through Extensive Reading' in Poedjosoedarmo, G. (ed.) *Innovative Approaches to Reading and Writing*. Singapore: RELC Anthology Series 46, pp.10-21.

Nation, I.S.P. (2009). 'Vocabulary size test' Available from: <http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 2009].

Nation, P. (2015). 'Principles guiding vocabulary learning through extensive reading' *Reading in a Foreign Language* 27 (1), pp.136–145.

Nation, I.S.P. & Beglar, D. (2007). 'A vocabulary size test' *The Language Teacher* 31 (7), pp.9-13.

Nation, I.S.P. & Coady, J. (1988). 'Vocabulary and reading' in Carter, R. & McCarthy, M. (eds.) *Vocabulary and Language Teaching*. London: Longman, pp.97-110.

Nation, I.S.P. & Wang Ming-tzu, K. (1999). 'Graded Readers and Vocabulary' *Reading in a Foreign Language* 12 (2), pp.355-380.

Nunan, D. (1992). *Research Methods in Language Learning*. Cambridge: Cambridge University Press.

O'Sullivan, A. (1993). 'Handbooks for teachers of English to Arab students' *ELT Journal* 47 (2), pp.172-175.

O'Sullivan, A. (2004). 'Reading and Arab College Students – Issues in the United Arab Emirates Higher Colleges of Technology' In *Proceedings of the First International Online Conference on Second and Foreign Language Teaching and Research: Beyond Borders*. Available from:

http://s3.amazonaws.com/academia.edu.documents/6190178/sullivan.pdf?AWSAccessKeyId=AKIAIWOWYYGZ2Y53UL3A&Expires=1490457543&Signature=X%2FoZITNEiOumNnsLcuWC%2Fmw7p0s%3D&response-content-disposition=inline%3B%20filename%3DReading_and_Arab_college_students_Issues.pdf [accessed 2016].

Pallant, J. (2010). *SPSS Survival Manual*. 4th Edition. Maidenhead: Open University Press McGraw-Hill.

Paran, A. (2008). 'The role of literature in instructed foreign language learning and teaching: An evidence-based survey' *Language Teaching* 41 (4), pp.465-496.

Pearson, A. (1968). 'Reflexions on the Use of Simplified Texts' *ELT Journal* 22 (3), pp.242-246.

Pigada, M. & Schmitt, N. (2006). 'Vocabulary acquisition from extensive reading: A case study' *Reading in a Foreign Language* 18 (1), pp.1-28.

Pellicer-Sanchez, A.M. & Schmitt, N. (2010). 'Incidental vocabulary acquisition from an authentic novel: Do *Things* Fall Apart?' *Reading in a Foreign Language* 22 (1), pp.31-55.

Poulshock, J. (2010). 'Extensive graded reading in the liberal arts and sciences' *Reading in a Foreign Language* 22 (2), pp.304-322.

Pritchard, R.M.O & Nasr, A. (2004). 'Improving reading performance among Egyptian engineering students: principles and practice' *English for Specific Purposes* 23 (4), pp.425-445.

Pulido, D. (2007). 'The Effects of Topic Familiarity and Passage Sight Vocabulary on L2 Lexical Inferencing and Retention through Reading' *Applied Linguistics* 28 (1), pp.66-86.

Reid-Thomas, H.C. & Hill, D.R. (1993). 'Survey review: Seventeen series of graded readers' *ELT Journal* 47 (3), pp.250-267.

Richards, K. (2003). *Qualitative Inquiry in TESOL*. Basingstoke: Palgrave Macmillan.

Ro, E. (2013). 'A case study of extensive reading with an unmotivated L2 reader' *Reading in a Foreign Language* 25 (2), pp.213–233.

Robb, T. & Kano, M. (2013). 'Effective extensive reading outside the classroom: A large-scale experiment' *Reading in a Foreign Language* 25 (2), pp.234–247.

Robinson, P. (1995). 'Attention, Memory, and the "Noticing" Hypothesis' *Language Learning* 45 (2), pp.283-331.

Ronnqvist, L. & Sell, R.D. (1994). 'Teenage books for teenagers: reflections on literature in language education' *ELT Journal* 48 (2), pp.125-132.

Rounds, P.L. (1992). 'Reading Aloud. Another Educator Comments' *TESOL Quarterly* 26 (4), pp.787-790.

Saragi, T., Nation, I.S.P. & Meister, G.F. (1978). 'Vocabulary Learning and Reading' *System* 6 (2), pp.72-78.

- Saville-Troike, M. (1973). 'Reading and the Audio-Lingual Method' *TESOL Quarterly* 7 (4), pp.395-405.
- Scanlon, C.L. (1942). 'The Purpose of Graded Texts' *The Modern Language Journal* 26 (6), pp.421-423.
- Schmidt, R. (1990). 'The Role of Consciousness in Second Language Learning' *Applied Linguistics* 11 (2), pp.129-158.
- Schmidt, R. (2012). 'Attention, awareness, and individual differences in language learning' in Chan, W.M., Chin, K.N., Bhatt, S.K. & Walker, I. (ed.) *Perspectives on Individual Characteristics and Foreign Language Education*. Boston/Berlin: Walter de Gruyter Inc., pp. 27-50.
- Schmitt, N. (1994). 'Vocabulary Testing: Questions for Test Development with Six Examples of Tests of Vocabulary Size and Depth' *Thai TESOL Bulletin* 6 (2), pp.9-16.
- Schmitt, N. (1997). 'Vocabulary learning strategies' In Schmitt, N. & McCarthy, M. (eds.), *Vocabulary: Description, Acquisition, and Pedagogy*. Cambridge: Cambridge University Press.
- Schmitt, N. (1998). 'Tracking the Incremental Acquisition of Second Language Vocabulary: A Longitudinal Study' *Language Learning* 48 (2), pp.281-317.
- Schmitt, N. (1999). 'Current Trends in Second Language Vocabulary' *TESOL Research Newsletter* 6 (2), pp.3-5.
- Schmitt, N., Jiang, X. & Grabe, W. (2011). 'The percentage of words known in a text and reading comprehension' *Modern Language Journal* 95 (1), pp.26-43.

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). 'Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test' *Language Testing* 18, pp.55-88.
- Sharwood Smith, M. (1981). 'Consciousness-Raising and the Second Language Learner' *Applied Linguistics* II (2), pp.159-168.
- Shaw, P., Irvine, A., Malmström, H. & Pecorari, D. (2010). 'Intertextual Episodes in Lectures: A Classification from the Perspective of Incidental Learning from Reading' *Journal of Language and Communication Studies* 45, pp.115-128.
- Sivell, J.N. (1980). 'Habitual Memorisation by Literature Students— a Help as Well as a Hindrance' *ELT Journal* 35 (1), pp.51-54.
- Spears, L.A. (1995). 'Nurses as technical writers: what they need to know' *Journal of Technical Writing and Communication* 25 (4), pp.401-414.
- Stoller, F.L. (2015). 'Viewing extensive reading from different vantage points' *Reading in a Foreign Language* 27 (1), pp.152–159.
- Strong, B.P. & Boutorwick, T.J. (2012). 'Incidental vocabulary acquisition from pleasure reading' *Extensive Reading World Congress Proceedings* 1, pp.71-74.
- Suk, N. (2016). *Impact of Extensive Reading in a Korean EFL University Setting: A Mixed Methods Study* (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (Accession Order No. 3708188). Available from: <http://pqdtopen.proquest.com/doc/1696319050.html?FMT=ABS&pubnum=3708188> [accessed 2016].
- Swann, J. (1994). 'Observing and Recording Talk in Educational Settings' In Graddol, D., Maybin, J. & Stierer, B. (eds.) *Researching Language and Literacy in Social Context*. Clevedon: Multilingual Matters Ltd, pp.26-48.

Tardy, C. (2004). 'The role of English in scientific communication: *lingua franca* or *Tyrannosaurus rex*?' *Journal of English for Academic Purposes* 3, pp.247-269.

Tien, C.Y. (2015). 'A large-scale study on extensive reading program for non-English majors: Factors and Attitudes' *International Journal of Applied Linguistics & English Literature* 4 (4), pp.46-54.

Vaismoradi, M., Turunen, H. & Bondas, T. (2013). 'Content analysis and thematic analysis: Implications for conducting a qualitative descriptive study' *Nursing and Health Sciences* 15, pp.398–405.

Wanarom, U. (2008). 'Comparing the vocabulary of different graded-reading schemes' *Reading in a Foreign Language* 20 (1), pp.43-69.

Waring, R. (2007). 'Bringing extensive reading into oral communication classes' *The Language Teacher* 31 (7), pp.38-39.

Waring, R. (2015). 'Exploration of the core and variable dimensions of extensive reading research and pedagogy' *Reading in a Foreign Language* 27 (1), pp.160-167.

Waring, R. & Nation, P. (2004). 'Second Language Reading and Incidental Vocabulary Learning' *Angles on the English-Speaking World* (4), pp.97-110.

Waring, R. & Takaki, M. (2003). 'At what rate do learners learn and retain new vocabulary from reading a graded reader?' *Reading in a Foreign Language* 15 (2), pp.130-163.

Webb, S. & Chang, A.C.-S. (2015). 'How does prior word knowledge affect vocabulary learning progress in an extensive reading program?' *Studies in Second Language Acquisition* 37 (4), pp.651-675.

West, M. (1964). 'Criteria in the Selection of Simplified Reading Books' *ELT Journal* 18 (4), pp.146-153.

Williams, R. (1986). '“Top ten” principles for teaching reading' *ELT Journal* 40 (1), pp.42-45.

Yamashita, J. (2004). 'Reading attitudes in L1 and L2, and their influence on L2 extensive reading' *Reading in a Foreign Language* 16 (1), pp.1-19.

Yamashita, J. (2015). 'In search of the nature of extensive reading in L2: Cognitive, affective, and pedagogical perspectives' *Reading in a Foreign Language* 27 (1), pp.168–181.

Yoshizawa, K. (2014). 'Edinburgh Project on Extensive Reading (EPER) Reading Comprehension Tests: Scoring and Setting Cutoff Scores' *Kansai University Bulletin of the Foreign Language Department* 10, pp.33-43.

Zhang, D. & Koda, K. (2012). 'Contribution of morphological awareness and lexical inferencing ability to L2 vocabulary knowledge and reading comprehension among advanced EFL learners: testing direct and indirect effects' *Reading and Writing* 25 (5), pp.1195-1216.

APPENDICES

LIST OF APPENDICES

Appendix-1	The Original Words in the Vocabulary Size Test (Nation, 2009) up to the 10000-word level [see Section-3.4.4]	p.259
Appendix-2	The Chosen Words for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) [see Section-3.4.4 for the selection criteria]	p.261
Appendix-3	Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2]	p.262
Appendix-4	Test-3 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.4.3]	p.266
Appendix-5	Summary of Specific Revisions made to Test-1 v.1 (Spelling) & Test-3 v.1 (Meaning) in light of the feedback during the Stage-1 Initial Pilot Study [see Section-3.4.4.3]	p.269
Appendix-6	Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1]	p.270
Appendix-7	Final List of Acceptable Translated Answers for Test-2 (Open Translation) during the Stage-3 Main Study [see Section-3.4.4.2]	p.274
Appendix-8	Test-3 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.3]	p.276
Appendix-9	Test-4 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.5]	p.279
Appendix-10	Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.5]	p.282
Appendix-11	Daily Reading Diary Sheet - Stage-1 Initial Pilot Study [see Section-3.6.1]	p.286
Appendix-12	Weekly Reading Diary Sheet - Stage-3 Main Study [see Section-3.6.1, Section-3.6.2 & Section-4.4]	p.287
Appendix-13	Survey of Exposure to English Learning - Stage-3 Main Study [see Section-3.7.1]	p.288
Appendix-14	Results of English Placement Test [see Section-4.1]	p.290
Appendix-15	Results of the Stage-2 Main Pilot Vocabulary Tests [see Section-3.4.7]	p.292
Appendix-16	Comparing Quantitative Data of the Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49) [see Section-3.9.1]	p.294
Appendix-17	In-Depth Interview Schedule [see Section-3.9.2.1]	p.297
Appendix-18	Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]	p.300
Appendix-19	Results of End-of-Semester English Course Exam [see Section-4.3]	p.305
Appendix-20	Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4]	p.307
Appendix-21	Results of the Survey of Exposure to English Learning [see Section-4.5.1]	p.314
Appendix-22	A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7]	p.320
Appendix-23	Initial Codes generated during Phase-2 of the Interview Analysis [see Section-4.7]	p.337
Appendix-24	Main Codes & Sub-Codes generated during Phase-3 of the Interview Analysis [see Section-4.7]	p.338
Appendix-25	Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7]	p.342
Appendix-26	Narratives and Themes refined during Phase-5 & Phase-6 of the Interview Analysis [see Section-4.7]	p.346

Appendix-1: The Original Words in the Vocabulary Size Test (Nation, 2009) up to the 10000-word level [see Section-3.4.4]

Word Frequency Level		Word with Example Sentence	Length	Word Class
1000-word level	1	see: They saw it.	3	verb
	2	time: They have a lot of time.	4	noun
	3	period: It was a difficult period.	6	noun
	4	figure: Is this the right figure?	6	noun
	5	poor: We are poor.	4	adjective (complement)
	6	drive: He drives fast.	5	verb
	7	jump: She tried to jump.	4	verb
	8	shoe: Where is your shoe?	4	noun
	9	standard: Her standards are very high.	8	noun
	10	basis: I don't understand the basis.	5	noun
2000-word level	1	maintain: Can they maintain it?	8	verb
	2	stone: He sat on a stone.	5	noun
	3	upset: I am upset.	5	adjective (complement)
	4	drawer: The drawer was empty.	6	noun
	5	patience: He has no patience.	8	noun
	6	nil: His mark for that question was nil.	3	adjective (complement)
	7	pub: They went to the pub.	3	noun
	8	circle: Make a circle.	6	noun
	9	microphone: Please use the microphone.	10	noun
	10	pro: He's a pro.	3	noun
3000-word level	1	soldier: He is a soldier.	7	noun
	2	restore: It has been restored.	7	verb
	3	jug: He was holding a jug.	3	noun
	4	scrub: He is scrubbing it.	5	verb
	5	dinosaur: The children were pretending to be dinosaurs.	8	noun
	6	strap: He broke the strap.	5	noun
	7	pave: It was paved.	4	verb
	8	dash: They dashed over it.	4	verb
	9	rove: He couldn't stop roving.	4	verb
	10	lonesome: He felt lonesome.	8	adjective (complement)
4000-word level	1	compound: They made a new compound.	8	noun
	2	latter: I agree with the latter.	6	noun
	3	candid: Please be candid.	6	adjective (complement)
	4	tummy: Look at my tummy.	5	noun
	5	quiz: We made a quiz.	4	noun
	6	input: We need more input.	5	noun
	7	crab: Do you like crabs?	4	noun
	8	vocabulary: You will need more vocabulary.	10	noun
	9	remedy: We found a good remedy.	6	noun
	10	allege: They alleged it.	6	verb
5000-word level	1	deficit: The company had a large deficit.	7	noun
	2	weep: He wept.	4	verb
	3	nun: We saw a nun.	3	noun
	4	haunt: The house is haunted.	5	verb
	5	compost: We need some compost.	7	noun
	6	cube: I need one more cube.	4	noun
	7	miniature: It is a miniature.	9	noun
	8	peel: Shall I peel it?	4	verb
	9	fracture: They found a fracture.	8	noun
	10	bacterium: They didn't find a single bacterium.	9	noun

Appendix-1: The Original Words in the Vocabulary Size Test (Nation, 2009) up to the 10000-word level [see Section-3.4.4] (cont.)

Word Frequency Level		Word with Example Sentence	Length	Word Class
6000-word level	1	devious: Your plans are devious.	7	adjective (complement)
	2	premier: The premier spoke for an hour.	7	noun
	3	butler: They have a butler.	6	noun
	4	accessory: They gave us some accessories.	9	noun
	5	threshold: They raised the threshold.	9	noun
	6	thesis: She has completed her thesis.	6	noun
	7	strangle: He strangled her.	8	verb
	8	cavalier: He treated her in a cavalier manner.	8	adjective
	9	malign: His malign influence is still felt.	6	adjective
	10	veer: The car veered.	4	verb
7000-word level	1	olive: We bought olives.	5	noun
	2	quilt: They made a quilt.	5	noun
	3	stealth: They did it by stealth.	7	noun
	4	shudder: The boy shuddered.	7	verb
	5	bristle: The bristles are too hard.	7	noun
	6	bloc: They have joined this bloc.	4	noun
	7	demography: This book is about demography.	10	noun
	8	gimmick: That's a good gimmick.	7	noun
	9	azalea: This azalea is very pretty.	6	noun
	10	yoghurt: This yoghurt is disgusting.	7	noun
8000-word level	1	erratic: He was erratic.	7	adjective (complement)
	2	palette: He lost his palette.	7	noun
	3	null: His influence was null.	4	adjective (complement)
	4	kindergarten: This is a good kindergarten.	10	noun
	5	eclipse: There was an eclipse.	7	noun
	6	marrow: This is the marrow.	6	noun
	7	locust: There were hundreds of locusts.	6	noun
	8	authentic: It is authentic.	9	adjective (complement)
	9	cabaret: We saw the cabaret.	7	noun
	10	mumble: He started to mumble.	6	verb
9000-word level	1	hallmark: Does it have a hallmark?	8	noun
	2	puritan: He is a puritan.	7	noun
	3	monologue: Now he has a monologue.	9	noun
	4	weir: We looked at the weir.	4	noun
	5	whim: He had lots of whims.	4	noun
	6	perturb: I was perturbed.	7	verb
	7	regent: They chose a regent.	6	noun
	8	octopus: They saw an octopus.	7	noun
	9	fen: The story is set in the fens.	3	noun
	10	lintel: He painted the lintel.	6	noun
10000-word level	1	awe: They looked at the mountain with awe.	3	noun
	2	peasantry: He did a lot for the peasantry.	9	noun
	3	egalitarian: This organization is very egalitarian.	11	adjective (complement)
	4	mystique: He has lost his mystique.	8	noun
	5	upbeat: I'm feeling really upbeat about it.	6	adjective (complement)
	6	cranny: We found it in the cranny!	6	noun
	7	pigtail: Does she have a pigtail?	7	noun
	8	crowbar: He used a crowbar.	7	noun
	9	ruck: He got hurt in the ruck.	4	noun
	10	lectern: He stood at the lectern.	7	noun

Appendix-2: The Chosen Words for Test-1 (Spelling), Test-2 (Open Translation) & Test-3 (Meaning) [see Section-3.4.4 for the selection criteria]

Word Frequency Level	Word with Example Sentence	Length	Word Class
1000-word level	1 see: They saw it.	3	verb
	2 time: They have a lot of time.	4	noun
	3 poor: We are poor.	4	adjective (complement)
	4 jump: She tried to jump.	4	verb
	5 basis: I don't understand the basis.	5	noun
2000-word level	6 maintain: Can they maintain it?	8	verb
	7 stone: He sat on a stone.	5	noun
	8 upset: I am upset.	5	adjective (complement)
	9 patience: He has no patience.	8	noun
	10 microphone: Please use the microphone.	10	noun
3000-word level	11 soldier: He is a soldier.	7	noun
	12 restore: It has been restored.	7	verb
	13 dinosaur: The children were pretending to be dinosaurs.	8	noun
	14 rove: He couldn't stop roving.	4	verb
	15 lonesome: He felt lonesome.	8	adjective (complement)
4000-word level	16 compound: They made a new compound.	8	noun
	17 latter: I agree with the latter.	6	noun
	18 candid: Please be candid.	6	adjective (complement)
	19 quiz: We made a quiz.	4	noun
	20 allege: They alleged it.	6	verb
5000-word level	21 deficit: The company had a large deficit.	7	noun
	22 weep: He wept.	4	verb
	23 haunt: The house is haunted.	5	verb
	24 miniature: It is a miniature.	9	noun
	25 peel: Shall I peel it?	4	verb
6000-word level	26 devious: Your plans are devious.	7	adjective (complement)
	27 premier: The premier spoke for an hour.	7	noun
	28 strangle: He strangled her.	8	verb
	29 malign: His malign influence is still felt.	6	adjective
	30 veer: The car veered.	4	verb
7000-word level	31 olive: We bought olives.	5	noun
	32 quilt: They made a quilt.	5	noun
	33 shudder: The boy shuddered.	7	verb
	34 demography: This book is about demography.	10	noun
	35 yoghurt: This yoghurt is disgusting.	7	noun
8000-word level	36 erratic: He was erratic.	7	adjective (complement)
	37 null: His influence was null.	4	adjective (complement)
	38 kindergarten: This is a good kindergarten.	10	noun
	39 eclipse: There was an eclipse.	7	noun
	40 mumble: He started to mumble.	6	verb
9000-word level	41 hallmark: Does it have a hallmark?	8	noun
	42 puritan: He is a puritan.	7	noun
	43 monologue: Now he has a monologue.	9	noun
	44 perturb: I was perturbed.	7	verb
	45 regent: They chose a regent.	6	noun
10000-word level	46 awe: They looked at the mountain with awe.	3	noun
	47 egalitarian: This organization is very egalitarian.	11	adjective (complement)
	48 mystique: He has lost his mystique.	8	noun
	49 upbeat: I'm feeling really upbeat about it.	6	adjective (complement)
	50 crowbar: He used a crowbar.	7	noun

Appendix-3: Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2]

Circle the letter [A, B, C or D] of the sentence which spells the bold underlined word in the correct way.

Write CLEARLY a word or sentence in ARABIC that explains the meaning of the English word

هات كلمة عربية أو جملة عربية تقتر معنى الكلمة الإنجليزية واكتبها بخط نسخ واضح

ضع دائرة حول تقييمك لدرجة التأكد من جوابك

مثلا: ضع دائرة حول "0" لو لم يكن لديك أية فكرة

مثلا: ضع دائرة حول "5" لو كنت متأكدًا جدًا

1	A They <u>siw</u> it.	B They <u>saw</u> it.	C They <u>sur</u> it.	D They <u>sor</u> it.	5	4	3	2	1	0
					5	4	3	2	1	0
2	A They have a lot of <u>tim</u> .	B They have a lot of <u>tyme</u> .	C They have a lot of <u>time</u> .	D They have a lot of <u>tym</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
3	A We are <u>poor</u> .	B We are <u>poar</u> .	C We are <u>por</u> .	D We are <u>pur</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
4	A She tried to <u>jimp</u> .	B She tried to <u>jamp</u> .	C She tried to <u>jump</u> .	D She tried to <u>jomp</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
5	A I don't understand the <u>baisis</u> .	B I don't understand the <u>baysis</u> .	C I don't understand the <u>basis</u> .	D I don't understand the <u>besis</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
6	A Can they <u>maintain</u> it?	B Can they <u>mantan</u> it?	C Can they <u>mantain</u> it?	D Can they <u>maintan</u> it?	5	4	3	2	1	0
					5	4	3	2	1	0
7	A He sat on a <u>stone</u> .	B He sat on a <u>stow</u> n.	C He sat on a <u>stoan</u> .	D He sat on a <u>stune</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
8	A I am <u>upsat</u> .	B I am <u>upset</u> .	C I am <u>upsit</u> .	D I am <u>upsot</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
9	A He has no <u>pashence</u> .	B He has no <u>pashunce</u> .	C He has no <u>patiance</u> .	D He has no <u>patience</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
10	A Please use the <u>mickrophone</u> .	B Please use the <u>microfone</u> .	C Please use the <u>mikrofone</u> .	D Please use the <u>microphone</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
11	A He is a <u>soljier</u> .	B He is a <u>soldier</u> .	C He is a <u>solqier</u> .	D He is a <u>solkier</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
12	A It has been <u>ristored</u> .	B It has been <u>restored</u> .	C It has been <u>restoared</u> .	D It has been <u>ristoared</u> .	5	4	3	2	1	0
					5	4	3	2	1	0

Appendix-3: Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2] (cont.)

13	A The children were pretending to be <u>dinosars</u> .	B The children were pretending to be <u>dinasaur</u> s.	C The children were pretending to be <u>dinasars</u> .	D The children were pretending to be <u>dinosaurs</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
14	A He couldn't stop <u>rowvinq</u> .	B He couldn't stop <u>rovwing</u> .	C He couldn't stop <u>roving</u> .	D He couldn't stop <u>roveing</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
15	A He felt <u>lonesome</u> .	B He felt <u>lonsome</u> .	C He felt <u>lonesum</u> .	D He felt <u>lonsum</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
16	A They made a new <u>compownd</u> .	B They made a new <u>compound</u> .	C They made a new <u>compawnd</u> .	D They made a new <u>compaund</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
17	A I agree with the <u>lattar</u> .	B I agree with the <u>latter</u> .	C I agree with the <u>latar</u> .	D I agree with the <u>lettar</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
18	A Please be <u>canded</u> .	B Please be <u>candud</u> .	C Please be <u>candad</u> .	D Please be <u>candid</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
19	A We made a <u>quaz</u> .	B We made a <u>quiz</u> .	C We made a <u>quez</u> .	D We made a <u>quoz</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
20	A They <u>alleged</u> it.	B They <u>alledged</u> it.	C They <u>aleged</u> it.	D They <u>aledged</u> it.	5	4	3	2	1	0
					5	4	3	2	1	0
21	A The company had a large <u>deficit</u> .	B The company had a large <u>defecit</u> .	C The company had a large <u>defecet</u> .	D The company had a large <u>deficet</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
22	A He <u>wept</u> .	B He <u>wapt</u> .	C He <u>wipt</u> .	D He <u>wopt</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
23	A The house is <u>hawnted</u> .	B The house is <u>hornted</u> .	C The house is <u>haunted</u> .	D The house is <u>hounted</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
24	A It is a <u>miniachure</u> .	B It is a <u>minichure</u> .	C It is a <u>miniature</u> .	D It is a <u>miniture</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
25	A Shall I <u>peel</u> it?	B Shall I <u>pele</u> it?	C Shall I <u>piel</u> it?	D Shall I <u>peil</u> it?	5	4	3	2	1	0
					5	4	3	2	1	0
26	A Your plans are <u>devious</u> .	B Your plans are <u>dewious</u> .	C Your plans are <u>devius</u> .	D Your plans are <u>dewius</u> .	5	4	3	2	1	0
					5	4	3	2	1	0

Appendix-3: Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2] (cont.)

27	A The <u>premiar</u> spoke for an hour.	B The <u>premier</u> spoke for an hour.	C The <u>pramier</u> spoke for an hour.	D The <u>pramiar</u> spoke for an hour.	5	4	3	2	1	0
					5	4	3	2	1	0
28	A He <u>straniled</u> her.	B He <u>strangeled</u> her.	C He <u>stranieled</u> her.	D He <u>strangled</u> her.	5	4	3	2	1	0
					5	4	3	2	1	0
29	A His <u>maline</u> influence is still felt.	B His <u>malign</u> influence is still felt.	C His <u>maliyn</u> influence is still felt.	D His <u>malene</u> influence is still felt.	5	4	3	2	1	0
					5	4	3	2	1	0
30	A The car <u>veared</u> .	B The car <u>weered</u> .	C The car <u>wearred</u> .	D The car <u>veered</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
31	A We bought <u>oliwes</u> .	B We bought <u>olives</u> .	C We bought <u>olivs</u> .	D We bought <u>oliws</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
32	A They made a <u>quelt</u> .	B They made a <u>qualt</u> .	C They made a <u>quolt</u> .	D They made a <u>guilt</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
33	A The boy <u>shudered</u> .	B The boy <u>shuddered</u> .	C The boy <u>shuderred</u> .	D The boy <u>shuddered</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
34	A This book is about <u>demografy</u> .	B This book is about <u>demografi</u> .	C This book is about <u>demography</u> .	D This book is about <u>demographi</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
35	A This <u>yoghurt</u> is disgusting.	B This <u>yogurt</u> is disgusting.	C This <u>yoghort</u> is disgusting.	D This <u>yogot</u> is disgusting.	5	4	3	2	1	0
					5	4	3	2	1	0
36	A He was <u>errattic</u> .	B He was <u>eratic</u> .	C He was <u>erattic</u> .	D He was <u>erratic</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
37	A His influence was <u>nill</u> .	B His influence was <u>nall</u> .	C His influence was <u>null</u> .	D His influence was <u>noll</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
38	A This is a good <u>kindergarten</u> .	B This is a good <u>kindergarden</u> .	C This is a good <u>cindergarten</u> .	D This is a good <u>cindergarden</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
39	A There was an <u>eklipse</u> .	B There was an <u>eclipse</u> .	C There was an <u>ecлипce</u> .	D There was an <u>eklipce</u> .	5	4	3	2	1	0
					5	4	3	2	1	0

Appendix-3: Test-1 v.1 (Spelling) & Test-2 v.1 (Open Translation) Combined - Stage-1 Initial Pilot Study [see Section-3.4.4.1 & Section-3.4.4.2] (cont.)

40	A He started to <u>mumbel</u> .	B He started to <u>mumbal</u> .	C He started to <u>mumble</u> .	D He started to <u>munble</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
41	A Does it have a <u>halmark</u> ?	B Does it have a <u>hallmark</u> ?	C Does it have a <u>hallmarc</u> ?	D Does it have a <u>halmarc</u> ?	5	4	3	2	1	0
					5	4	3	2	1	0
42	A He is a <u>puritan</u> .	B He is a <u>puretan</u> .	C He is a <u>poritan</u> .	D He is a <u>poretan</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
43	A Now he has a <u>monolog</u> .	B Now he has a <u>monologe</u> .	C Now he has a <u>monologue</u> .	D Now he has a <u>monolojue</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
44	A I was <u>preturbed</u> .	B I was <u>parturbed</u> .	C I was <u>priturbed</u> .	D I was <u>perturbed</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
45	A They chose a <u>reagent</u> .	B They chose a <u>regent</u> .	C They chose a <u>reegent</u> .	D They chose a <u>regant</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
46	A They looked at the mountain with <u>orwe</u> .	B They looked at the mountain with <u>arwe</u> .	C They looked at the mountain with <u>awe</u> .	D They looked at the mountain with <u>uwe</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
47	A This organization is very <u>egalitarien</u> .	B This organization is very <u>aqalitarian</u> .	C This organization is very <u>egalitarian</u> .	D This organization is very <u>aqalitarien</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
48	A He has lost his <u>mystiq</u> .	B He has lost his <u>mystique</u> .	C He has lost his <u>mystik</u> .	D He has lost his <u>mysteague</u> .	5	4	3	2	1	0
					5	4	3	2	1	0
49	A I'm feeling really <u>upbeat</u> about it.	B I'm feeling really <u>upbeet</u> about it.	C I'm feeling really <u>upbeit</u> about it.	D I'm feeling really <u>upbiet</u> about it.	5	4	3	2	1	0
					5	4	3	2	1	0
50	A He used a <u>croebar</u> .	B He used a <u>crowebar</u> .	C He used a <u>crowbare</u> .	D He used a <u>crowbar</u> .	5	4	3	2	1	0
					5	4	3	2	1	0

Appendix-4: Test-3 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.4.3]

Circle the letter [A, B, C or D] which explains the meaning of the word in the most correct way.

Circle the number that indicates how sure you were of your answer

ضع دائرة حول تقييمك لدرجة التأكد من جوابك

E.g. Circle '0' if you have no idea and are just blindly guessing

مثلاً: ضع دائرة حول "0" لو لم يكن لديك أية فكرة

E.g. Circle '5' if you feel absolutely sure

مثلاً: ضع دائرة حول "5" لو كنت متأكدًا جدًا

1 see: They saw it.	a. cut	b. waited for	c. looked at	d. started	5 4 3 2 1 0
2 time: They have a lot of time.	a. money	b. food	c. hours	d. friends	5 4 3 2 1 0
3 poor: We are poor.	a. have no money	b. feel happy	c. are very interested	d. do not like to work hard	5 4 3 2 1 0
4 jump: She tried to jump.	a. lie on top of the water	b. get off the ground suddenly	c. stop the car at the edge of the road	d. move very fast	5 4 3 2 1 0
5 basis: I don't understand the basis.	a. reason	b. words	c. road signs	d. main part	5 4 3 2 1 0
6 maintain: Can they maintain it?	a. keep it as it is	b. make it larger	c. get a better one than it	d. get it	5 4 3 2 1 0
7 stone: He sat on a stone.	a. hard thing	b. kind of chair	c. soft thing on the floor	d. part of a tree	5 4 3 2 1 0
8 upset: I am upset.	a. tired	b. famous	c. rich	d. unhappy	5 4 3 2 1 0
9 patience: He has no patience.	a. will not wait happily	b. has no free time	c. has no faith	d. does not know what is fair	5 4 3 2 1 0
10 microphone: Please use the microphone.	a. machine for making food hot	b. machine that makes sounds louder	c. machine that makes things look bigger	d. small telephone that can be carried around	5 4 3 2 1 0
11 soldier: He is a soldier.	a. person in a business	b. student	c. person who uses metal	d. person in the army	5 4 3 2 1 0
12 restore: It has been restored.	a. said again	b. given to a different person	c. given a lower price	d. made like new again	5 4 3 2 1 0
13 dinosaur: The children were pretending to be dinosaurs.	a. robbers who work at sea	b. very small creatures with human form but with wings	c. large creatures with wings that breathe fire	d. animals that lived an extremely long time ago	5 4 3 2 1 0
14 rove: He couldn't stop roving.	a. getting drunk	b. traveling around	c. making a musical sound through closed lips	d. working hard	5 4 3 2 1 0
15 lonesome: He felt lonesome.	a. ungrateful	b. very tired	c. lonely	d. full of energy	5 4 3 2 1 0
16 compound: They made a new compound.	a. agreement	b. thing made of two or more parts	c. group of people forming a business	d. guess based on past experience	5 4 3 2 1 0
17 latter: I agree with the latter.	a. man from the church	b. reason given	c. last one	d. answer	5 4 3 2 1 0
18 candid: Please be candid.	a. be careful	b. show sympathy	c. show fairness to both sides	d. say what you really think	5 4 3 2 1 0
19 quiz: We made a quiz.	a. thing to hold arrows	b. serious mistake	c. set of questions	d. box for birds to make nests in	5 4 3 2 1 0
20 allege: They alleged it.	a. claimed it without proof	b. stole the ideas for it from someone else	c. provided facts to prove it	d. argued against the facts that supported it	5 4 3 2 1 0
21 deficit: The company had a large deficit.	a. spent a lot more money than it earned	b. went down a lot in value	c. had a plan for its spending that used a lot of money	d. had a lot of money stored in the bank	5 4 3 2 1 0
22 weep: He wept.	a. finished his course	b. cried	c. died	d. worried	5 4 3 2 1 0

Appendix-4: Test-3 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.4.3]

(cont.)

23 haunt: The house is haunted.	a. full of ornaments	b. rented	c. empty	d. full of ghosts	5 4 3 2 1 0
24 miniature: It is a miniature.	a. a very small thing of its kind	b. an instrument for looking at very small objects	c. a very small living creature	d. a small line to join letters in handwriting	5 4 3 2 1 0
25 peel: Shall I peel it?	a. let it sit in water for a long time	b. take the skin off it	c. make it white	d. cut it into thin pieces	5 4 3 2 1 0
26 devious: Your plans are devious.	a. tricky	b. well-developed	c. not well thought out	d. more expensive than necessary	5 4 3 2 1 0
27 premier: The premier spoke for an hour.	a. person who works in a law court	b. university teacher	c. adventurer	d. head of the government	5 4 3 2 1 0
28 strangle: He strangled her.	a. killed her by pressing her throat	b. gave her all the things she wanted	c. took her away by force	d. admired her greatly	5 4 3 2 1 0
29 malign: His malign influence is still felt.	a. evil	b. good	c. very important	d. secret	5 4 3 2 1 0
30 veer: The car veered.	a. went suddenly in another direction	b. moved shakily	c. made a very loud noise	d. slid sideways without the wheels turning	5 4 3 2 1 0
31 olive: We bought olives.	a. oily fruit	b. scented pink or red flowers	c. men's clothes for swimming	d. tools for digging up weeds	5 4 3 2 1 0
32 quilt: They made a quilt.	a. statement about who should get their property when they die	b. firm agreement	c. thick warm cover for a bed	d. feather pen	5 4 3 2 1 0
33 shudder: The boy shuddered.	a. spoke with a low voice	b. almost fell	c. shook	d. called out loudly	5 4 3 2 1 0
34 demography: This book is about demography.	a. the study of patterns of land use	b. the study of the use of pictures to show facts about numbers	c. the study of the movement of water	d. the study of population	5 4 3 2 1 0
35 yoghurt: This yoghurt is disgusting.	a. dark grey mud found at the bottom of rivers	b. unhealthy, open sore	c. thick, soured milk, often with sugar and flavouring	d. large purple fruit with soft flesh	5 4 3 2 1 0
36 erratic: He was erratic.	a. without fault	b. very bad	c. very polite	d. unsteady	5 4 3 2 1 0
37 null: His influence was null.	a. had good results	b. was unhelpful	c. had no effect	d. was long-lasting	5 4 3 2 1 0
38 kindergarten: This is a good kindergarten.	a. activity that allows you to forget your worries	b. place of learning for children too young for school	c. strong, deep bag carried on the back	d. place where you may borrow books	5 4 3 2 1 0
39 eclipse: There was an eclipse.	a. a strong wind	b. a loud noise of something hitting the water	c. the killing of a large number of people	d. the sun hidden by a planet	5 4 3 2 1 0
40 mumble: He started to mumble.	a. think deeply	b. shake uncontrollably	c. stay further behind the others	d. speak in an unclear way	5 4 3 2 1 0
41 hallmark: Does it have a hallmark?	a. stamp to show when it should be used by	b. stamp to show the quality	c. mark to show it is approved by the royal family	d. mark or stain to prevent copying	5 4 3 2 1 0
42 puritan: He is a puritan.	a. person who likes attention	b. person with strict morals	c. person with a moving home	d. person who keeps money and hates spending it	5 4 3 2 1 0
43 monologue: Now he has a monologue.	a. single piece of glass to hold over his eye to help him to see better	b. long turn at talking without being interrupted	c. position with all the power	d. picture made by joining letters together in interesting ways	5 4 3 2 1 0

Appendix-4: Test-3 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.4.3]
(cont.)

44 perturb: I was perturbed.	a. made to agree	b. worried	c. very puzzled	d. very wet	5 4 3 2 1 0
45 regent: They chose a regent.	a. an irresponsible person	b. a person to run a meeting for a short time	c. a ruler acting in place of the king or queen	d. a person to represent them	5 4 3 2 1 0
46 awe: They looked at the mountain with awe.	a. worry	b. interest	c. wonder	d. respect	5 4 3 2 1 0
47 egalitarian: This organization is very egalitarian.	a. does not provide much information about itself to the public	b. dislikes change	c. frequently asks a court of law for a judgement	d. treats everyone who works for it as if they are equal	5 4 3 2 1 0
48 mystique: He has lost his mystique.	a. his healthy body	b. the secret way he makes other people think he has special power or skill	c. the woman who has been his lover while he is married to someone else	d. the hair on his top lip	5 4 3 2 1 0
49 upbeat: I'm feeling really upbeat about it.	a. upset	b. good	c. hurt	d. confused	5 4 3 2 1 0
50 crowbar: He used a crowbar.	a. heavy iron pole with a curved end	b. false name	c. sharp tool for making holes in leather	d. light metal walking stick	5 4 3 2 1 0

Appendix-5: Summary of Specific Revisions made to Test-1 v.1 (Spelling) & Test-3 v.1 (Meaning) in light of the feedback during the Stage-1 Initial Pilot Study [see Section-3.4.4.3]

Test	Word Knowledge	Item	Word and Example Sentence	Option	Original	Revision in light of the feedback during the Stage 1 Initial Pilot Study	Reason
Test-1	Spelling	35		A	yogurt	yogort	Original is another correct spelling
		37		A	nill	nell	Original is the correct spelling of another word with a different meaning
Test-3	Meaning	5	basis: I don't understand the basis	A	reason	last part	Original deemed to be too close in meaning to the correct option i.e. D 'main part'
		15	lonesome: He felt lonesome	C	lonely	without a friend	Original deemed to be too obvious, i.e. too similar to the item word
		20	allege: They alleged it	C	provided facts to prove it	hit it with a stick	Original deemed to be too close in meaning to the correct option i.e. A 'claimed it without proof'
		44	perturb: I was perturbed	B	worried	very content	Original deemed to be too close in meaning to the correct option i.e. C 'very puzzled'
		46	awe: They looked at the mountain with awe	D	respect	disrespect	Original deemed to be too close in meaning to the correct option i.e. C 'wonder'
		49	upbeat: I'm feeling really upbeat about it	B	good	happy	Original deemed to be unclear

Appendix-6: Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1]

Circle the letter [A, B, C or D] of the sentence which spells the bold underlined word in the correct way.

Write CLEARLY a word or sentence in ARABIC that explains the meaning of the English word

هات كلمة عربية أو جملة عربية تفسر معنى الكلمة الإنجليزية الصحيحة واكتبها بخط نسخ واضح

Circle the number that indicates how sure you were of your answer

ضع دائرة حول تقييمك لدرجة التأكد من جوابك

E.g. Circle '0' if you have no idea and are just blindly guessing

مثلاً: ضع دائرة حول "0" لو لم يكن لديك أية فكرة

E.g. Circle '3' if you feel absolutely sure

مثلاً: ضع دائرة حول "3" لو كنت متأكدًا جدًا

1	A They <u>siw</u> it.	B They <u>saw</u> it.	C They <u>sur</u> it.	D They <u>sor</u> it.	3 2 1 0
					3 2 1 0
2	A They have a lot of <u>tim</u> .	B They have a lot of <u>tyne</u> .	C They have a lot of <u>time</u> .	D They have a lot of <u>tym</u> .	3 2 1 0
					3 2 1 0
3	A We are <u>poor</u> .	B We are <u>poar</u> .	C We are <u>por</u> .	D We are <u>pur</u> .	3 2 1 0
					3 2 1 0
4	A She tried to <u>jimp</u> .	B She tried to <u>jamp</u> .	C She tried to <u>jump</u> .	D She tried to <u>jomp</u> .	3 2 1 0
					3 2 1 0
5	A I don't understand the <u>baisis</u> .	B I don't understand the <u>baysis</u> .	C I don't understand the <u>basis</u> .	D I don't understand the <u>besis</u> .	3 2 1 0
					3 2 1 0
6	A Can they <u>maintain</u> it?	B Can they <u>mantan</u> it?	C Can they <u>mantain</u> it?	D Can they <u>maintan</u> it?	3 2 1 0
					3 2 1 0
7	A He sat on a <u>stone</u> .	B He sat on a <u>stown</u> .	C He sat on a <u>stoan</u> .	D He sat on a <u>stune</u> .	3 2 1 0
					3 2 1 0
8	A I am <u>upsat</u> .	B I am <u>upset</u> .	C I am <u>upsit</u> .	D I am <u>upsof</u> .	3 2 1 0
					3 2 1 0
9	A He has no <u>pashence</u> .	B He has no <u>pashunce</u> .	C He has no <u>patiance</u> .	D He has no <u>patience</u> .	3 2 1 0
					3 2 1 0
10	A Please use the <u>mickrophone</u> .	B Please use the <u>microfone</u> .	C Please use the <u>mikrofone</u> .	D Please use the <u>microphone</u> .	3 2 1 0
					3 2 1 0
11	A He is a <u>soljier</u> .	B He is a <u>soldier</u> .	C He is a <u>solqier</u> .	D He is a <u>solkier</u> .	3 2 1 0
					3 2 1 0
12	A It has been <u>ristored</u> .	B It has been <u>restored</u> .	C It has been <u>restoared</u> .	D It has been <u>ristoared</u> .	3 2 1 0
					3 2 1 0

Appendix-6: Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1] (cont.)

13	A The children were pretending to be <u>dinosars</u> .	B The children were pretending to be <u>dinasaurs</u> .	C The children were pretending to be <u>dinasars</u> .	D The children were pretending to be <u>dinosaurs</u> .	3 2 1 0
					3 2 1 0
14	A He couldn't stop <u>rowving</u> .	B He couldn't stop <u>rovwing</u> .	C He couldn't stop <u>roving</u> .	D He couldn't stop <u>roveing</u> .	3 2 1 0
					3 2 1 0
15	A He felt <u>lonesome</u> .	B He felt <u>lonsome</u> .	C He felt <u>lonesum</u> .	D He felt <u>lonsum</u> .	3 2 1 0
					3 2 1 0
16	A They made a new <u>compownd</u> .	B They made a new <u>compound</u> .	C They made a new <u>compawnd</u> .	D They made a new <u>compaund</u> .	3 2 1 0
					3 2 1 0
17	A I agree with the <u>lattar</u> .	B I agree with the <u>latter</u> .	C I agree with the <u>latar</u> .	D I agree with the <u>lettat</u> .	3 2 1 0
					3 2 1 0
18	A Please be <u>canded</u> .	B Please be <u>candud</u> .	C Please be <u>candad</u> .	D Please be <u>candid</u> .	3 2 1 0
					3 2 1 0
19	A We made a <u>quaz</u> .	B We made a <u>quiz</u> .	C We made a <u>quez</u> .	D We made a <u>quoz</u> .	3 2 1 0
					3 2 1 0
20	A They <u>alleged</u> it.	B They <u>alledged</u> it.	C They <u>aleged</u> it.	D They <u>aledged</u> it.	3 2 1 0
					3 2 1 0
21	A The company had a large <u>deficit</u> .	B The company had a large <u>defecit</u> .	C The company had a large <u>defecet</u> .	D The company had a large <u>deficet</u> .	3 2 1 0
					3 2 1 0
22	A He <u>wept</u> .	B He <u>wapt</u> .	C He <u>wipt</u> .	D He <u>wopt</u> .	3 2 1 0
					3 2 1 0
23	A The house is <u>hawnted</u> .	B The house is <u>hornted</u> .	C The house is <u>haunted</u> .	D The house is <u>hounted</u> .	3 2 1 0
					3 2 1 0
24	A It is a <u>miniachure</u> .	B It is a <u>minichure</u> .	C It is a <u>miniature</u> .	D It is a <u>miniture</u> .	3 2 1 0
					3 2 1 0
25	A Shall I <u>pee</u> it?	B Shall I <u>pele</u> it?	C Shall I <u>pie</u> it?	D Shall I <u>peil</u> it?	3 2 1 0
					3 2 1 0
26	A Your plans are <u>devious</u> .	B Your plans are <u>dewious</u> .	C Your plans are <u>devius</u> .	D Your plans are <u>dewius</u> .	3 2 1 0
					3 2 1 0

Appendix-6: Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1] (cont.)

27	A The <u>premiar</u> spoke for an hour.	B The <u>premier</u> spoke for an hour.	C The <u>pramier</u> spoke for an hour.	D The <u>pramiar</u> spoke for an hour.	3 2 1 0
					3 2 1 0
28	A He <u>stranjied</u> her.	B He <u>strangeled</u> her.	C He <u>stranieled</u> her.	D He <u>strangled</u> her.	3 2 1 0
					3 2 1 0
29	A His <u>maline</u> influence is still felt.	B His <u>malign</u> influence is still felt.	C His <u>maliyn</u> influence is still felt.	D His <u>malene</u> influence is still felt.	3 2 1 0
					3 2 1 0
30	A The car <u>veared</u> .	B The car <u>weered</u> .	C The car <u>weared</u> .	D The car <u>veered</u> .	3 2 1 0
					3 2 1 0
31	A We bought <u>oliwes</u> .	B We bought <u>olives</u> .	C We bought <u>olivs</u> .	D We bought <u>oliws</u> .	3 2 1 0
					3 2 1 0
32	A They made a <u>guelt</u> .	B They made a <u>qualt</u> .	C They made a <u>quolt</u> .	D They made a <u>guilt</u> .	3 2 1 0
					3 2 1 0
33	A The boy <u>shudered</u> .	B The boy <u>shudderred</u> .	C The boy <u>shuderred</u> .	D The boy <u>shuddered</u> .	3 2 1 0
					3 2 1 0
34	A This book is about <u>demogrfy</u> .	B This book is about <u>demogرافي</u> .	C This book is about <u>demography</u> .	D This book is about <u>demographi</u> .	3 2 1 0
					3 2 1 0
35	A This <u>yoghurt</u> is disgusting.	B This <u>yogort</u> is disgusting.	C This <u>yoghort</u> is disgusting.	D This <u>yogot</u> is disgusting.	3 2 1 0
					3 2 1 0
36	A He was <u>errattic</u> .	B He was <u>eratic</u> .	C He was <u>erattic</u> .	D He was <u>erratic</u> .	3 2 1 0
					3 2 1 0
37	A His influence was <u>nell</u> .	B His influence was <u>nall</u> .	C His influence was <u>null</u> .	D His influence was <u>noll</u> .	3 2 1 0
					3 2 1 0
38	A This is a good <u>kindergarten</u> .	B This is a good <u>kindergarden</u> .	C This is a good <u>cinderqarten</u> .	D This is a good <u>cinderqarden</u> .	3 2 1 0
					3 2 1 0
39	A There was an <u>eklipse</u> .	B There was an <u>eclipse</u> .	C There was an <u>ecclipce</u> .	D There was an <u>eklipce</u> .	3 2 1 0
					3 2 1 0
40	A He started to <u>mumbel</u> .	B He started to <u>mumbal</u> .	C He started to <u>mumble</u> .	D He started to <u>munble</u> .	3 2 1 0
					3 2 1 0

Appendix-6: Test-1 v.2 (Spelling) & Test-2 v.2 (Open Translation) Combined - Stage-2 Main Pilot Study & Stage-3 Main Study [see Section-3.4.4.1] (cont.)

41	A Does it have a <u>halmark?</u>	B Does it have a <u>hallmark?</u>	C Does it have a <u>hallmarc?</u>	D Does it have a <u>halmarc?</u>	3 2 1 0
					3 2 1 0
42	A He is a <u>puritan.</u>	B He is a <u>puretan.</u>	C He is a <u>poritan.</u>	D He is a <u>poretan.</u>	3 2 1 0
					3 2 1 0
43	A Now he has a <u>monolog.</u>	B Now he has a <u>monologe.</u>	C Now he has a <u>monologue.</u>	D Now he has a <u>monolojue.</u>	3 2 1 0
					3 2 1 0
44	A I was <u>preturbed.</u>	B I was <u>parturbed.</u>	C I was <u>priturbed.</u>	D I was <u>perturbed.</u>	3 2 1 0
					3 2 1 0
45	A They chose a <u>reagent.</u>	B They chose a <u>regent.</u>	C They chose a <u>reegent.</u>	D They chose a <u>regant.</u>	3 2 1 0
					3 2 1 0
46	A They looked at the mountain with <u>orwe.</u>	B They looked at the mountain with <u>arwe.</u>	C They looked at the mountain with <u>awe.</u>	D They looked at the mountain with <u>uwe.</u>	3 2 1 0
					3 2 1 0
47	A This organization is very <u>egalitarien.</u>	B This organization is very <u>agaltarian.</u>	C This organization is very <u>egalitarian.</u>	D This organization is very <u>agalitarien.</u>	3 2 1 0
					3 2 1 0
48	A He has lost his <u>mystig.</u>	B He has lost his <u>mystique.</u>	C He has lost his <u>mystik.</u>	D He has lost his <u>mysteague.</u>	3 2 1 0
					3 2 1 0
49	A I'm feeling really <u>upbeat</u> about it.	B I'm feeling really <u>upbeet</u> about it.	C I'm feeling really <u>upbeit</u> about it.	D I'm feeling really <u>upbiet</u> about it.	3 2 1 0
					3 2 1 0
50	A He used a <u>croebar.</u>	B He used a <u>crowebar.</u>	C He used a <u>crowbare.</u>	D He used a <u>crowbar.</u>	3 2 1 0
					3 2 1 0

Appendix-7: Final List of Acceptable Translated Answers for Test-2 (Open Translation) during the Stage-3 Main Study [see Section-3.4.4.2]

	Words from Test-2 (Open Translation)	Close but Unacceptable Translations	Acceptable Translations
1	see: They saw it.		ينظر - شافوا - شاهدوا رأوا
2	time: They have a lot of time.		دقيقة - الوقت
3	poor: We are poor.		ضعيف - فقراء - مسكين
4	jump: She tried to jump.	NOT climb تسلق	الفقر - جيمار - تمارين
5	basis: I don't understand the basis.		قواعد - الأساسيات - قاعدة - أساس - مبدىء
6	maintain: Can they maintain it?	NOT بقاء	صيانة - إصلاح - يثبتوا يحافظوا عليها - تصليح
7	stone: He sat on a stone.		حجر - صخر
8	upset: I am upset.		مرهق - زعلان - غاضب - قلق - معصب - حزين - شعور بالحزن - متزعج - محيط - مستاء - متضايق - ضجر - مضطجر
9	patience: He has no patience.		صبر - حليم
10	microphone: Please use the microphone.		مايك - ميكروفون سماعة - لاقط صوت - مكبر الصوت - مائكروفون - تلفون
11	soldier: He is a soldier.		ضابط - عسكري - حربي - محارب - جندي - قائد - فارس - شرطي
12	restore: It has been restored.		أعيد - أعيدت - أرجعت - استرجاع - استعيدت - أصلحت - رمت - إعادة التصنيع - مسترجع - جدد - إعادته
13	dinosaur: The children were pretending to be dinosaurs.		ديناصورات
14	rove: He couldn't stop roving.		الحركة - يتحرك - الجري و الهرولة - يتجول - يتطوف
15	lonesome: He felt lonesome.		وحدة - وحيدا - يعيش بمفرده
16	compound: They made a new compound.	NOT تكوين	سكن - مركب - مجمع سكني - خليط - عناصر - عنصر - مصنع - تركيبة - نوع من المنازل - مسكن
17	latter: I agree with the latter.		الأخر - الأخير - الثاني - الأخرى - الباقي
18	candid: Please be candid.		صريحا - صريح
19	quiz: We made a quiz.		اختبار قصير أو بسيط - نوع اختبار - تجربة - امتحان - سؤال - مسابقة - تقييم
20	allege: They alleged it.	NOT falsify, invent, contrive لفقوا	ادعى - زعم
21	deficit: The company had a large deficit.	NOT نقص	عجز مالي
22	weep: He wept.		بكوا - يتوج
23	haunt: The house is haunted.		مسكون - مسكون بالجن - مسكون بالاشباح - مطارد بالاشباح - مسكون بالتيماطين
24	miniature: It is a miniature.		تقلص - صورة مصغرة - مجسم مصغر - صغير - نسخة مصغرة - رسم مصغر
25	peel: Shall I peel it?		أقشر - كشط
26	devious: Your plans are devious.		شريرة - مكررة ملتوية - مخادع
27	premier: The premier spoke for an hour.	NOT ممثل as in premier league, or مندوب	أولي - مدير - مقدم - رئيس الوزراء - الحكومة - عمدة
28	strangle: He strangled her.	NOT ربطها	ختفها
29	malign: His malign influence is still felt.		شديد - مؤذي
30	veer: The car veered.		حولت مسارها - انحرفت بسرعة - انحرفت
31	olive: We bought olives.		زيتون - زيت
32	quilt: They made a quilt.		لحاف - بطانية
33	shudder: The boy shuddered.		ارتعش - تردد
34	demography: This book is about demography.		جغرافيا - علم السكان - التوصيف أو التوزيع السكاني - الديموغرافي - احصائية السكان
35	yoghurt: This yoghurt is disgusting.		لين روب - لين رائب - زيادي - لين
36	erratic: He was erratic.	NOT مخطأ	غير منظم - متقلب المزاج - غريب الأطوار - شاذ
37	null: His influence was null.		فارغ - لا شيء - باطل - صفر - لا يذكر - قاضي - خالي

Appendix-7: Final List of Acceptable Translated Answers for Test-2 (Open Translation) during the Stage-3 Main Study [see Section-3.4.4.2] (cont.)

	Words from Test-2 (Open Translation)	Close but Unacceptable Translations	Acceptable Translations
38	kindergarten: This is a good kindergarten.		ملاهي - حديقة أطفال - تمهيدى - روضة الأطفال
39	eclipse: There was an eclipse.		خسوف - كسوف
40	mumble: He started to mumble.	NOT ثرثرة	يتكلم دون وضوح - يتلعثم - stutter - stammer - كلام غير مفهوم - يتكلم بصوت خافت - تمتمة - كلام غير واضح
41	hallmark: Does it have a hallmark?		درجة كاملة - ميزة - علامة فارقة - علامة الجودة - لوحة الدرجة - درجة عالمية - درجة مرتفعة - كامل الدرجة - علامة مرتفعة - علامة مميزة
42	puritan: He is a puritan.		متقشف - aesthetic life - ok - تطهري - متشدد - مغالي - متزمت
43	monologue: Now he has a monologue.	NOT حوار داخلي - مع النفس	حوار ذاتي - مناجاة فردية
44	perturb: I was perturbed.		مضطرب - انزعجت - أقلق
45	regent: They chose a regent.		تائب - بديل - وصي العرش
46	awe: They looked at the mountain with awe.		انبهار - خوف - خشية - خشوع - يعجب - هلع
47	egalitarian: This organization is very egalitarian.		منصفة - عدالي - متساوي
48	mystique: He has lost his mystique.		غموض - غامض - سحر - خاصية التموض و الغازية - charm - mystery
49	upbeat: I'm feeling really upbeat about it.		جيد - متحمس - متفائل - مرتاح اليال - يارتياح
50	crowbar: He used a crowbar.		رافعة - lifting equipment - عصا - أداة لفتح الصناديق

Appendix-8: Test-3 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.4.3]

Circle the letter [A, B, C or D] which explains the meaning of the word in the most correct way.

Circle the number that indicates how sure you were of your answer

E.g. Circle '0' if you have no idea and are just blindly guessing

E.g. Circle '3' if you feel absolutely sure

ضع دائرة حول تقييمك لدرجة التأكد من جوابك

مثلا: ضع دائرة حول "0" لو لم يكن لديك أية فكرة

مثلا: ضع دائرة حول "3" لو كنت متأكدًا جدًا

1 see: They saw it.	a. cut	b. waited for	c. looked at	d. started	3 2 1 0
2 time: They have a lot of time.	a. money	b. food	c. hours	d. friends	3 2 1 0
3 poor: We are poor.	a. have no money	b. feel happy	c. are very interested	d. do not like to work hard	3 2 1 0
4 jump: She tried to jump.	a. lie on top of the water	b. get off the ground suddenly	c. stop the car at the edge of the road	d. move very fast	3 2 1 0
5 basis: I don't understand the basis.	a. last part	b. words	c. road signs	d. main part	3 2 1 0
6 maintain: Can they maintain it?	a. keep it as it is	b. make it larger	c. get a better one than it	d. get it	3 2 1 0
7 stone: He sat on a stone.	a. hard thing	b. kind of chair	c. soft thing on the floor	d. part of a tree	3 2 1 0
8 upset: I am upset.	a. tired	b. famous	c. rich	d. unhappy	3 2 1 0
9 patience: He has no patience.	a. will not wait happily	b. has no free time	c. has no faith	d. does not know what is fair	3 2 1 0
10 microphone: Please use the microphone.	a. machine for making food hot	b. machine that makes sounds louder	c. machine that makes things look bigger	d. small telephone that can be carried around	3 2 1 0
11 soldier: He is a soldier.	a. person in a business	b. student	c. person who uses metal	d. person in the army	3 2 1 0
12 restore: It has been restored.	a. said again	b. given to a different person	c. given a lower price	d. made like new again	3 2 1 0
13 dinosaur: The children were pretending to be dinosaurs.	a. robbers who work at sea	b. very small creatures with human form but with wings	c. large creatures with wings that breathe fire	d. animals that lived an extremely long time ago	3 2 1 0
14 rove: He couldn't stop roving.	a. getting drunk	b. traveling around	c. making a musical sound through closed lips	d. working hard	3 2 1 0
15 lonesome: He felt lonesome.	a. ungrateful	b. very tired	c. without a friend	d. full of energy	3 2 1 0
16 compound: They made a new compound.	a. agreement	b. thing made of two or more parts	c. group of people forming a business	d. guess based on past experience	3 2 1 0
17 latter: I agree with the latter.	a. man from the church	b. reason given	c. last one	d. answer	3 2 1 0
18 candid: Please be candid.	a. be careful	b. show sympathy	c. show fairness to both sides	d. say what you really think	3 2 1 0
19 quiz: We made a quiz.	a. thing to hold arrows	b. serious mistake	c. set of questions	d. box for birds to make nests in	3 2 1 0
20 allege: They alleged it.	a. claimed it without proof	b. stole the ideas for it from someone else	c. hit it with a stick	d. argued against the facts that supported it	3 2 1 0
21 deficit: The company had a large deficit.	a. spent a lot more money than it earned	b. went down a lot in value	c. had a plan for its spending that used a lot of money	d. had a lot of money stored in the bank	3 2 1 0
22 weep: He wept.	a. finished his course	b. cried	c. died	d. worried	3 2 1 0

Appendix-8: Test-3 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.4.3] (cont.)

23	haunt: The house is haunted.	a. full of ornaments	b. rented	c. empty	d. full of ghosts	3	2	1	0
24	miniature: It is a miniature.	a. a very small thing of its kind	b. an instrument for looking at very small objects	c. a very small living creature	d. a small line to join letters in handwriting	3	2	1	0
25	peel: Shall I peel it?	a. let it sit in water for a long time	b. take the skin off it	c. make it white	d. cut it into thin pieces	3	2	1	0
26	devious: Your plans are devious.	a. tricky	b. well-developed	c. not well thought out	d. more expensive than necessary	3	2	1	0
27	premier: The premier spoke for an hour.	a. person who works in a law court	b. university teacher	c. adventurer	d. head of the government	3	2	1	0
28	strangle: He strangled her.	a. killed her by pressing her throat	b. gave her all the things she wanted	c. took her away by force	d. admired her greatly	3	2	1	0
29	malign: His malign influence is still felt.	a. evil	b. good	c. very important	d. secret	3	2	1	0
30	veer: The car veered.	a. went suddenly in another direction	b. moved shakily	c. made a very loud noise	d. slid sideways without the wheels turning	3	2	1	0
31	olive: We bought olives.	a. oily fruit	b. scented pink or red flowers	c. men's clothes for swimming	d. tools for digging up weeds	3	2	1	0
32	quilt: They made a quilt.	a. statement about who should get their property when they die	b. firm agreement	c. thick warm cover for a bed	d. feather pen	3	2	1	0
33	shudder: The boy shuddered.	a. spoke with a low voice	b. almost fell	c. shook	d. called out loudly	3	2	1	0
34	demography: This book is about demography.	a. the study of patterns of land use	b. the study of the use of pictures to show facts about numbers	c. the study of the movement of water	d. the study of population	3	2	1	0
35	yoghurt: This yoghurt is disgusting.	a. dark grey mud found at the bottom of rivers	b. unhealthy, open sore	c. thick, soured milk, often with sugar and flavouring	d. large purple fruit with soft flesh	3	2	1	0
36	erratic: He was erratic.	a. without fault	b. very bad	c. very polite	d. unsteady	3	2	1	0
37	null: His influence was null.	a. had good results	b. was unhelpful	c. had no effect	d. was long-lasting	3	2	1	0
38	kindergarten: This is a good kindergarten.	a. activity that allows you to forget your worries	b. place of learning for children too young for school	c. strong, deep bag carried on the back	d. place where you may borrow books	3	2	1	0
39	eclipse: There was an eclipse.	a. a strong wind	b. a loud noise of something hitting the water	c. the killing of a large number of people	d. the sun hidden by a planet	3	2	1	0
40	mumble: He started to mumble.	a. think deeply	b. shake uncontrollably	c. stay further behind the others	d. speak in an unclear way	3	2	1	0
41	hallmark: Does it have a hallmark?	a. stamp to show when it should be used by	b. stamp to show the quality	c. mark to show it is approved by the royal family	d. mark or stain to prevent copying	3	2	1	0
42	puritan: He is a puritan.	a. person who likes attention	b. person with strict morals	c. person with a moving home	d. person who keeps money and hates spending it	3	2	1	0
43	monologue: Now he has a monologue.	a. single piece of glass to hold over his eye to help him to see better	b. long turn at talking without being interrupted	c. position with all the power	d. picture made by joining letters together in interesting ways	3	2	1	0

Appendix-8: Test-3 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.4.3] (cont.)

44 perturb: I was perturbed.	a. made to agree	b. very content	c. very puzzled	d. very wet	3 2 1 0
45 regent: They chose a regent.	a. an irresponsible person	b. a person to run a meeting for a short time	c. a ruler acting in place of the king or queen	d. a person to represent them	3 2 1 0
46 awe: They looked at the mountain with awe.	a. worry	b. interest	c. wonder	d. disrespect	3 2 1 0
47 egalitarian: This organization is very egalitarian.	a. does not provide much information about itself to the public	b. dislikes change	c. frequently asks a court of law for a judgement	d. treats everyone who works for it as if they are equal	3 2 1 0
48 mystique: He has lost his mystique.	a. his healthy body	b. the secret way he makes other people think he has special power or skill	c. the woman who has been his lover while he is married to someone else	d. the hair on his top lip	3 2 1 0
49 upbeat: I'm feeling really upbeat about it.	a. upset	b. happy	c. hurt	d. confused	3 2 1 0
50 crowbar: He used a crowbar.	a. heavy iron pole with a curved end	b. false name	c. sharp tool for making holes in leather	d. light metal walking stick	3 2 1 0

1 coffee	1 money for	1 arrange	1 grow
2 disease	2 work	2 develop	2 put in order
3 justice	3 a piece of	3 lean	3 like more than
4 skirt	4 clothing	4 owe	4 something else
5 stage	5 using the law	5 prefer	5 else
6 wage	6 in the right way	6 seize	6

1 clerk	1 a drink	1 blame	1 make
2 frame	2 office worker	2 elect	2 choose by
3 noise	3 unwanted	3 jump	3 voting
4 respect	4 sound	4 threaten	4 become like
5 theater	5	5 melt	5 water
6 wine	6	6 manufacture	6

1 dozen	1 chance	1 ancient	1 not easy
2 empire	2 twelve	2 curious	2 very old
3 gift	3 money paid	3 difficult	3 related to God
4 tax	4 to the	4 entire	4
5 relief	5 government	5 holy	5
6 opportunity	6	6 social	6

1 admire	1 make wider or	1 slight	1 beautiful
2 complain	2 longer	2 bitter	2 small
3 fix	3 bring in for	3 lovely	3 liked by many
4 hire	4 the first time	4 merry	4 people
5 introduce	5 have a high	5 popular	5
6 stretch	6 opinion of someone	6 independent	6

1 bull	1 formal and	1 muscle	1 advice
2 champion	2 serious	2 counsel	2 a place
3 dignity	3 manner	3 factor	3 covered with
4 hell	4 winner of a	4 hen	4 grass
5 museum	5 sporting event	5 lawn	5 female
6 solution	6 building where valuable objects are shown	6 atmosphere	6 chicken

Copyright © 2007 by Norbert Schmitt, Diane Schmitt and Caroline Clapham. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Appendix 1 Student instruction sheet for the Levels Test

This is a vocabulary test. You must choose the right word to go with each meaning. Write the number of that word next to its meaning. Here is an example.

1 business	1 part of a house
2 clock	2 animal with four legs
3 horse	3 something used for writing
4 pencil	4
5 shoe	5
6 wall	6

You answer it in the following way.

1 business → 6
2 clock → 1
3 horse → 2
4 pencil → 3
5 shoe → 4
6 wall → 5

10" SHOWS THE "6" WAS A PURE GUESS
5" SHOWS THE "3" WAS CERTAIN
2" SHOWS THE "4" WAS NOT CERTAIN BUT BETTER THAN A PURE GUESS

Some words are in the test to make it more difficult. You do not have to find a meaning for these words. In the example above, these words are *business, clock and shoe*.

If you have no idea about the meaning of a word, do not guess. But if you think you might know the meaning, then you should try to find the answer.

Appendix 2 The Vocabulary Levels Test: Version 2

(© Norbert Schmitt)

The 2000 word level

1 copy	1 end or highest	1 accident	1 loud deep
2 event	2 point	2 debt	2 sound
3 motor	3 this moves a	3 fortune	3 something you
4 pity	4 car	4 pride	4 must pay
5 profit	5 thing made to	5 roar	5 having a high
6 tip	6 be like another	6 thread	6 opinion of yourself

Copyright © 2007 by Norbert Schmitt, Diane Schmitt and Caroline Clapham. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording, or by any information storage and retrieval system, without permission in writing from the publisher.

Appendix-9: Test-4 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.5]
(cont.)

84 Two versions of the Vocabulary Levels Test		Norbert Schmitt, Diane Schmitt and Caroline Clapham 85	
1 blanket	— holiday	1 abandon	— live in a place
2 contest	— good quality	2 dwell	— follow in
3 generation	— wool covering	3 oblige	— order to catch
4 merit	— used on	4 pursue	— leave
5 plot	— beds	5 quote	— something permanently
6 vacation	—	6 resolve	—
1 comment	— long formal	1 assemble	— look closely
2 gown	— dress	2 attach	— stop doing
3 import	— goods from a	3 peer	— something
4 nerve	— foreign	4 quit	— cry out loudly
5 pasture	— country	5 scream	— in fear
6 tradition	— part of the body which carries feeling	6 loss	—
1 pond	— group of animals	1 drift	— suffer
2 angel	— spirit who serves God	2 endure	— patiently
3 frost	—	3 grasp	— join wool threads
4 herd	— managing business and affairs	4 knit	— together
5 fort	—	5 register	— hold firmly with your hands
6 administration	—	6 tumble	—
1 brilliant	— thin	1 aware	— usual
2 distinct	— steady	2 blank	— best or most important
3 magic	— without clothes	3 desperate	— knowing what is happening
4 naked	—	4 normal	—
5 slender	—	5 striking	—
6 stable	—	6 supreme	—
Academic Vocabulary		The 5000 word level	
1 area	— written	1 analysis	— eagerness
2 contract	— agreement	2 curb	— loan to buy a
3 definition	— way of doing	3 gravel	— house
4 evidence	— something	4 mortgage	— small
5 method	— reason for	5 scar	— stones
6 role	— believing something is or is not true	6 zeal	— mixed with sand
1 debate	— plan	1 access	— male or female
2 exposure	— choice	2 gender	— study of the mind
3 integration	— joining	3 psychology	—
4 option	— something into a whole	4 license	— entrance or
5 scheme	—	5 oration	— way in
6 stability	—	6 implementation	—
1 alter	— change	1 success	— correspond
2 coincide	— say something is not true	2 diminish	— keep
3 deny	— describe clearly and exactly	3 emerge	— match or be in agreement with
4 devote	—	4 highlight	—
5 release	—	5 invoke	— give special attention to something
6 specify	—	6 retain	—
1 bond	— make smaller	1 edition	— collecting things over time
2 channel	— guess the number or size of something	2 accumulation	—
3 estimate	— recognizing and naming a person or thing	3 guarantee	—
4 identify	—	4 media	— promise to repair a broken product
5 mediate	—	5 motivation	—
6 minimize	—	6 phenomenon	—
1 abstract	— next to	1 explicit	— something
2 adjacent	— added to	2 final	— last
3 neutral	— concerning the whole world	3 negative	— stiff
4 global	—	4 professional	— meaning 'no' or 'not'
5 controversial	—	5 rigid	—
6 supplementary	—	6 sole	—
1 artillery	— a kind of tree	1 analysis	— eagerness
2 creed	— system of belief	2 curb	— loan to buy a
3 hydrogen	—	3 gravel	— house
4 maple	—	4 mortgage	— small
5 pork	— large gun	5 scar	— stones
6 streak	— on wheels	6 zeal	— mixed with sand

Appendix-9: Test-4 v.1 (Meaning) - Stage-1 Initial Pilot Study [see Section-3.4.5]
(cont.)

The 10 000 word level

1 alabaster	1 small barrel	1 throttle	1 kindness
2 tentacle	2 soft white	2 convey	2 set of musical
3 stigma	3 stone	3 lien	3 notes
4 keg	4 tool for	4 octave	4 speed control
5 rasp	5 shaping wood	5 stink	5 for an
6 chandelier	6	6 benevolence	6 engine
1 bourgeois	1 middle class	1 scrawl	1 write
2 brocade	2 people	2 cringe	2 carelessly
3 consonant	3 row or level	3 immerse	3 move back
4 prelude	4 of something	4 peck	4 because of
5 stupor	5 cloth with a	5 contaminate	5 fear
6 tier	6 pattern or	6 relay	6 put something
	gold or silver		under water
	threads		
1 alcove	1 priest	1 blurt	1 walk in a
2 impectus	2 release from	2 dabble	2 proud way
3 maggot	3 prison early	3 dent	3 kill by
4 parole	4 medicine to	4 pacify	4 squeezing
5 salve	5 put on	5 strangle	5 someone's
6 vicar	6 wounds	6 swagger	6 throat
			say suddenly
			without
			thinking
1 alkali	1 light joking	1 illicit	1 immense
2 butter	2 talk	2 lewd	2 against the
3 coop	3 a rank of	3 mammoth	3 law
4 mosaic	4 British	4 slick	4 wanting
5 stealth	5 nobility	5 temporal	5 revenge
6 viscount	6 picture made	6 vindictive	
	of small pieces		
	of glass or		
	stone		
1 dissipate	1 steal	1 indolent	1 lazy
2 flaunt	2 scatter or	2 nocturnal	2 no longer
3 impede	3 vanish	3 obsolete	3 used
4 loot	4 twist the	4 torrid	4 clever and
5 squirm	5 body about	5 translucent	5 tricky
6 vie	6 uncomfortably	6 witty	

Downloaded from <http://ml.sagepub.com> at UNIV OF SOUTHERN CALIFORNIA on March 3, 2015
© 2015 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

1 cavalry	1 small hill	1 chart	1 nap
2 eve	2 day or night	2 forge	2 large beautiful
3 ham	3 before a	3 mansion	3 house
4 mound	4 holiday	4 outfit	4 place where
5 steak	5 soldiers who	5 sample	5 metals are
6 switch	6 fight from	6 volunteer	6 made and
	horses		shaped
1 circus	1 musical	1 revive	1 think about
2 jungle	2 instrument	2 extract	2 deeply
3 trumpet	3 seat without	3 gamble	3 bring back to
4 sermon	4 a back or	4 launch	4 health
5 stool	5 arms	5 provoke	5 make
6 nomination	6 speech	6 contemplate	6 someone
	given by a		angry
	priest in a		
	church		
1 shatter	1 have a rest	1 decent	1 weak
2 embarrass	2 break	2 frail	2 concerning a
3 heave	3 suddenly into	3 harsh	3 city
4 obscure	4 small	4 incredible	4 difficult to
5 demonstrate	5 pieces	5 municipal	5 believe
6 relax	6 make	6 specific	
	someone feel		
	shy or		
	nervous		
1 correspond	1 exchange	1 adequate	1 enough
2 embroider	2 letters	2 internal	2 fully grown
3 lurk	3 hide and wait	3 mature	3 alone away
4 penetrate	4 for someone	4 profound	4 from other
5 prescribe	5 feel angry	5 solitary	5 things
6 resent	6 about	6 tragic	
	something		

Downloaded from <http://ml.sagepub.com> at UNIV OF SOUTHERN CALIFORNIA on March 3, 2015
© 2015 SAGE Publications. All rights reserved. Not for commercial use or unauthorized distribution.

Appendix-10: Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.5]

ضع دائرة حول الكلمة الصحيحة.
Circle the letter [A, B, C, D, E or F] of the word which explains the sentence on the left in most correct way.
ضع دائرة حول تقييمك لدرجة التأكد من جوابك
مثلاً: ضع دائرة حول "0" لو لم يكن لديك أية فكرة
مثلاً: ضع دائرة حول "3" لو كنت متأكدًا جدًا

2000 Word Level							
1 end or highest point	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
2 this moves a car	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
3 thing made to be like another	A copy	B event	C motor	D pity	E profit	F tip	3 2 1 0
4 loud deep sound	A accident	B debt	C fortune	D pride	E roar	F thread	3 2 1 0
5 something you must pay	A accident	B debt	C fortune	D pride	E roar	F thread	3 2 1 0
6 having a high opinion of yourself	A accident	B debt	C fortune	D pride	E roar	F thread	3 2 1 0
7 money for work	A coffee	B disease	C justice	D skirt	E stage	F wage	3 2 1 0
8 a piece of clothing	A coffee	B disease	C justice	D skirt	E stage	F wage	3 2 1 0
9 using the law in the right way	A coffee	B disease	C justice	D skirt	E stage	F wage	3 2 1 0
10 grow	A arrange	B develop	C lean	D owe	E prefer	F seize	3 2 1 0
11 put in order	A arrange	B develop	C lean	D owe	E prefer	F seize	3 2 1 0
12 like more than something else	A arrange	B develop	C lean	D owe	E prefer	F seize	3 2 1 0
13 a drink	A clerk	B frame	C noise	D respect	E theater	F wine	3 2 1 0
14 office worker	A clerk	B frame	C noise	D respect	E theater	F wine	3 2 1 0
15 unwanted sound	A clerk	B frame	C noise	D respect	E theater	F wine	3 2 1 0
16 make	A blame	B elect	C jump	D threaten	E melt	F manufacture	3 2 1 0
17 choose by voting	A blame	B elect	C jump	D threaten	E melt	F manufacture	3 2 1 0
18 become like water	A blame	B elect	C jump	D threaten	E melt	F manufacture	3 2 1 0
19 chance	A dozen	B empire	C gift	D tax	E relief	F opportunity	3 2 1 0
20 twelve	A dozen	B empire	C gift	D tax	E relief	F opportunity	3 2 1 0
21 money paid to the government	A dozen	B empire	C gift	D tax	E relief	F opportunity	3 2 1 0
22 not easy	A ancient	B curious	C difficult	D entire	E holy	F social	3 2 1 0
23 very old	A ancient	B curious	C difficult	D entire	E holy	F social	3 2 1 0
24 related to God	A ancient	B curious	C difficult	D entire	E holy	F social	3 2 1 0
25 make wider or longer	A admire	B complain	C fix	D hire	E introduce	F stretch	3 2 1 0
26 bring in for the first time	A admire	B complain	C fix	D hire	E introduce	F stretch	3 2 1 0
27 have a high opinion of someone	A admire	B complain	C fix	D hire	E introduce	F stretch	3 2 1 0
28 beautiful	A slight	B bitter	C lovely	D merry	E popular	F independent	3 2 1 0
29 small	A slight	B bitter	C lovely	D merry	E popular	F independent	3 2 1 0
30 liked by many people	A slight	B bitter	C lovely	D merry	E popular	F independent	3 2 1 0

3000 Word Level							
1 formal and serious manner	A bull	B champion	C dignity	D hell	E museum	F solution	3 2 1 0
2 winner of a sporting event	A bull	B champion	C dignity	D hell	E museum	F solution	3 2 1 0
3 building where valuable objects are shown	A bull	B champion	C dignity	D hell	E museum	F solution	3 2 1 0
4 advice	A muscle	B counsel	C factor	D hen	E lawn	F atmosphere	3 2 1 0
5 a place covered with grass	A muscle	B counsel	C factor	D hen	E lawn	F atmosphere	3 2 1 0
6 female chicken	A muscle	B counsel	C factor	D hen	E lawn	F atmosphere	3 2 1 0
7 holiday	A blanket	B contest	C generation	D merit	E plot	F vacation	3 2 1 0
8 good quality	A blanket	B contest	C generation	D merit	E plot	F vacation	3 2 1 0
9 wool covering used on beds	A blanket	B contest	C generation	D merit	E plot	F vacation	3 2 1 0
10 live in a place	A abandon	B dwell	C oblige	D pursue	E quote	F resolve	3 2 1 0

Appendix-10: Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.5] (cont.)

11 follow in order to catch	A abandon	B dwell	C oblige	D pursue	E quote	F resolve	3 2 1 0
12 leave something permanently	A abandon	B dwell	C oblige	D pursue	E quote	F resolve	3 2 1 0
13 long formal dress	A comment	B gown	C import	D nerve	E pasture	F tradition	3 2 1 0
14 goods from a foreign country	A comment	B gown	C import	D nerve	E pasture	F tradition	3 2 1 0
15 part of the body which carries feeling	A comment	B gown	C import	D nerve	E pasture	F tradition	3 2 1 0
16 look closely	A assemble	B attach	C peer	D quit	E scream	F toss	3 2 1 0
17 stop doing something	A assemble	B attach	C peer	D quit	E scream	F toss	3 2 1 0
18 cry out loudly in fear	A assemble	B attach	C peer	D quit	E scream	F toss	3 2 1 0
19 group of animals	A pond	B angel	C frost	D herd	E fort	F administration	3 2 1 0
20 spirit who serves God	A pond	B angel	C frost	D herd	E fort	F administration	3 2 1 0
21 managing business and affairs	A pond	B angel	C frost	D herd	E fort	F administration	3 2 1 0
22 suffer patiently	A drift	B endure	C grasp	D knit	E register	F tumble	3 2 1 0
23 join wool threads together	A drift	B endure	C grasp	D knit	E register	F tumble	3 2 1 0
24 hold firmly with your hands	A drift	B endure	C grasp	D knit	E register	F tumble	3 2 1 0
25 thin	A brilliant	B distinct	C magic	D naked	E slender	F stable	3 2 1 0
26 steady	A brilliant	B distinct	C magic	D naked	E slender	F stable	3 2 1 0
27 without clothes	A brilliant	B distinct	C magic	D naked	E slender	F stable	3 2 1 0
28 usual	A aware	B blank	C desperate	D normal	E striking	F supreme	3 2 1 0
29 best or most important	A aware	B blank	C desperate	D normal	E striking	F supreme	3 2 1 0
30 knowing what is happening	A aware	B blank	C desperate	D normal	E striking	F supreme	3 2 1 0

Academic Word Level

1 written agreement	A area	B contract	C definition	D evidence	E method	F role	3 2 1 0
2 way of doing something	A area	B contract	C definition	D evidence	E method	F role	3 2 1 0
3 reason for believing something is or is not true	A area	B contract	C definition	D evidence	E method	F role	3 2 1 0
4 end	A adult	B vehicle	C exploitation	D infrastructure	E termination	F schedule	3 2 1 0
5 machine used to move people or goods	A adult	B vehicle	C exploitation	D infrastructure	E termination	F schedule	3 2 1 0
6 list of things to do at certain times	A adult	B vehicle	C exploitation	D infrastructure	E termination	F schedule	3 2 1 0
7 plan	A debate	B exposure	C integration	D option	E scheme	F stability	3 2 1 0
8 choice	A debate	B exposure	C integration	D option	E scheme	F stability	3 2 1 0
9 joining something into a whole	A debate	B exposure	C integration	D option	E scheme	F stability	3 2 1 0
10 change	A alter	B coincide	C deny	D devote	E release	F specify	3 2 1 0
11 say something is not true	A alter	B coincide	C deny	D devote	E release	F specify	3 2 1 0
12 describe clearly and exactly	A alter	B coincide	C deny	D devote	E release	F specify	3 2 1 0
13 male or female	A access	B gender	C psychology	D license	E orientation	F implementation	3 2 1 0
14 study of the mind	A access	B gender	C psychology	D license	E orientation	F implementation	3 2 1 0
15 entrance or way in	A access	B gender	C psychology	D license	E orientation	F implementation	3 2 1 0
16 keep	A correspond	B diminish	C emerge	D highlight	E invoke	F retain	3 2 1 0
17 match or be in agreement with	A correspond	B diminish	C emerge	D highlight	E invoke	F retain	3 2 1 0
18 give special attention to something	A correspond	B diminish	C emerge	D highlight	E invoke	F retain	3 2 1 0
19 collecting things over time	A edition	B accumulation	C guarantee	D media	E motivation	F phenomenon	3 2 1 0
20 promise to repair a broken product	A edition	B accumulation	C guarantee	D media	E motivation	F phenomenon	3 2 1 0

Appendix-10: Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.5] (cont.)

21	feeling a strong reason or need to do something	A edition	B accumulation	C guarantee	D media	E motivation	F phenomenon	3	2	1	0
22	make smaller	A bond	B channel	C estimate	D identify	E mediate	F minimize	3	2	1	0
23	guess the number or size of something	A bond	B channel	C estimate	D identify	E mediate	F minimize	3	2	1	0
24	recognizing and naming a person or thing	A bond	B channel	C estimate	D identify	E mediate	F minimize	3	2	1	0
25	last	A explicit	B final	C negative	D professional	E rigid	F sole	3	2	1	0
26	stiff	A explicit	B final	C negative	D professional	E rigid	F sole	3	2	1	0
27	meaning 'no' or 'not'	A explicit	B final	C negative	D professional	E rigid	F sole	3	2	1	0
28	next to	A abstract	B adjacent	C neutral	D global	E controversial	F supplementary	3	2	1	0
29	added to	A abstract	B adjacent	C neutral	D global	E controversial	F supplementary	3	2	1	0
30	concerning the whole world	A abstract	B adjacent	C neutral	D global	E controversial	F supplementary	3	2	1	0

5000 Word Level

1	eagerness	A analysis	B curb	C gravel	D mortgage	E scar	F zeal	3	2	1	0
2	loan to buy a house	A analysis	B curb	C gravel	D mortgage	E scar	F zeal	3	2	1	0
3	small stones mixed with sand	A analysis	B curb	C gravel	D mortgage	E scar	F zeal	3	2	1	0
4	a kind of tree	A artillery	B creed	C hydrogen	D maple	E pork	F streak	3	2	1	0
5	system of belief	A artillery	B creed	C hydrogen	D maple	E pork	F streak	3	2	1	0
6	large gun on wheels	A artillery	B creed	C hydrogen	D maple	E pork	F streak	3	2	1	0
7	small hill	A cavalry	B eve	C ham	D mound	E steak	F switch	3	2	1	0
8	day or night before a holiday	A cavalry	B eve	C ham	D mound	E steak	F switch	3	2	1	0
9	soldiers who fight from horses	A cavalry	B eve	C ham	D mound	E steak	F switch	3	2	1	0
10	map	A chart	B forge	C mansion	D outfit	E sample	F volunteer	3	2	1	0
11	large beautiful house	A chart	B forge	C mansion	D outfit	E sample	F volunteer	3	2	1	0
12	place where metals are made and shaped	A chart	B forge	C mansion	D outfit	E sample	F volunteer	3	2	1	0
13	musical instrument	A circus	B jungle	C trumpet	D sermon	E stool	F nomination	3	2	1	0
14	seat without a back or arms	A circus	B jungle	C trumpet	D sermon	E stool	F nomination	3	2	1	0
15	speech given by a priest in a church	A circus	B jungle	C trumpet	D sermon	E stool	F nomination	3	2	1	0
16	think about deeply	A revive	B extract	C gamble	D launch	E provoke	F contemplate	3	2	1	0
17	bring back to health	A revive	B extract	C gamble	D launch	E provoke	F contemplate	3	2	1	0
18	make someone angry	A revive	B extract	C gamble	D launch	E provoke	F contemplate	3	2	1	0
19	have a rest	A shatter	B embarrass	C heave	D obscure	E demonstrate	F relax	3	2	1	0
20	break suddenly into small pieces	A shatter	B embarrass	C heave	D obscure	E demonstrate	F relax	3	2	1	0
21	make someone feel shy or nervous	A shatter	B embarrass	C heave	D obscure	E demonstrate	F relax	3	2	1	0
22	weak	A decent	B frail	C harsh	D incredible	E municipal	F specific	3	2	1	0
23	concerning a city	A decent	B frail	C harsh	D incredible	E municipal	F specific	3	2	1	0
24	difficult to believe	A decent	B frail	C harsh	D incredible	E municipal	F specific	3	2	1	0
25	exchange letters	A correspond	B embroider	C lurk	D penetrate	E prescribe	F resent	3	2	1	0
26	hide and wait for someone	A correspond	B embroider	C lurk	D penetrate	E prescribe	F resent	3	2	1	0
27	feel angry about something	A correspond	B embroider	C lurk	D penetrate	E prescribe	F resent	3	2	1	0
28	enough	A adequate	B internal	C mature	D profound	E solitary	F tragic	3	2	1	0
29	fully grown	A adequate	B internal	C mature	D profound	E solitary	F tragic	3	2	1	0
30	alone away from other things	A adequate	B internal	C mature	D profound	E solitary	F tragic	3	2	1	0

Appendix-10: Test-4 v.2 (Meaning) - Stage-2 Main Pilot Study & Stage-3 Main Study
[see Section-3.4.5] (cont.)

10000 Word Level							
1 small barrel	A alabaster	B tentacle	C dogma	D keg	E rasp	F chandelier	3 2 1 0
2 soft white stone	A alabaster	B tentacle	C dogma	D keg	E rasp	F chandelier	3 2 1 0
3 tool for shaping wood	A alabaster	B tentacle	C dogma	D keg	E rasp	F chandelier	3 2 1 0
4 kindness	A throttle	B convoy	C lien	D octave	E stint	F benevolence	3 2 1 0
5 set of musical notes	A throttle	B convoy	C lien	D octave	E stint	F benevolence	3 2 1 0
6 speed control for an engine	A throttle	B convoy	C lien	D octave	E stint	F benevolence	3 2 1 0
7 middle class people	A bourgeois	B brocade	C consonant	D prelude	E stupor	F tier	3 2 1 0
8 row or level of something	A bourgeois	B brocade	C consonant	D prelude	E stupor	F tier	3 2 1 0
9 cloth with a pattern of gold or silver threads	A bourgeois	B brocade	C consonant	D prelude	E stupor	F tier	3 2 1 0
10 write carelessly	A scrawl	B cringe	C immerse	D peek	E contaminate	F relay	3 2 1 0
11 move back because of fear	A scrawl	B cringe	C immerse	D peek	E contaminate	F relay	3 2 1 0
12 put something under water	A scrawl	B cringe	C immerse	D peek	E contaminate	F relay	3 2 1 0
13 priest	A alcove	B impetus	C maggot	D parole	E salve	F vicar	3 2 1 0
14 release from prison early	A alcove	B impetus	C maggot	D parole	E salve	F vicar	3 2 1 0
15 medicine to put on wounds	A alcove	B impetus	C maggot	D parole	E salve	F vicar	3 2 1 0
16 walk in a proud way	A blurt	B dabble	C dent	D pacify	E strangle	F swagger	3 2 1 0
17 kill by squeezing someone's throat	A blurt	B dabble	C dent	D pacify	E strangle	F swagger	3 2 1 0
18 say suddenly without thinking	A blurt	B dabble	C dent	D pacify	E strangle	F swagger	3 2 1 0
19 light joking talk	A alkali	B banter	C coop	D mosaic	E stealth	F viscount	3 2 1 0
20 a rank of British nobility	A alkali	B banter	C coop	D mosaic	E stealth	F viscount	3 2 1 0
21 picture made of small pieces of glass or stone	A alkali	B banter	C coop	D mosaic	E stealth	F viscount	3 2 1 0
22 immense	A illicit	B lewd	C mammoth	D slick	E temporal	F vindictive	3 2 1 0
23 against the law	A illicit	B lewd	C mammoth	D slick	E temporal	F vindictive	3 2 1 0
24 wanting revenge	A illicit	B lewd	C mammoth	D slick	E temporal	F vindictive	3 2 1 0
25 steal	A dissipate	B flaunt	C impede	D loot	E squirm	F vie	3 2 1 0
26 scatter or vanish	A dissipate	B flaunt	C impede	D loot	E squirm	F vie	3 2 1 0
27 twist the body about uncomfortably	A dissipate	B flaunt	C impede	D loot	E squirm	F vie	3 2 1 0
28 lazy	A indolent	B nocturnal	C obsolete	D torrid	E translucent	F wily	3 2 1 0
29 no longer used	A indolent	B nocturnal	C obsolete	D torrid	E translucent	F wily	3 2 1 0
30 clever and tricky	A indolent	B nocturnal	C obsolete	D torrid	E translucent	F wily	3 2 1 0

Appendix-11: Daily Reading Diary Sheet - Stage-1 Initial Pilot Study [see Section-3.6.1]

1. Write the DAY when the reading took place:	يوم القراءة:						
2. Write the DATE when the reading took place:	تاريخ القراءة:						
3. Specify the LANGUAGE of the reading material:	ENGLISH	OR	أو	اللغة العربية	حدد لغة المادة المقروءة:		
4. Describe the TYPE of reading material:	حدد نوع المادة المقروءة:						
5. Write the TITLE or brief description:	اكتب عنوان المادة المقروءة أو وصفها الموجز:						
6. WHEN did you read the material?	ضع دائرة حول أوقات قراءة هذه المادة:						
	12am-6am	6am-8am	8am-12pm	12pm-4pm	4pm-9pm	9pm-12am	ليلاً صباحاً
7. HOW MUCH TIME in total did you spent reading the material?	ضع دائرة حول مدة قراءة هذه المادة الإجمالية:						
	Under 10min	10-20min	20-40min	40min-1hour	1-2hours	Over 2hours	أكثر من ساعتين أقل من 10 دقيقة
8. WHY did you read the material?	حدد أسباب قراءة هذه المادة:						
9. WHERE did you read the material?	حدد مكان قراءة هذه المادة:						
10. HOW EASY was the material to read?	ضع دائرة حول تقييمك لسهولة قراءة هذه المادة:						
	Very easy	5	4	3	2	1	Very hard
	سهلة جداً						صعبة جداً
11. HOW OFTEN did you use the DICTIONARY or glossary?	في استخدام القاموس خلال القراءة						
	>5wds/pg	2-5 wds/pg	1 wd/pg	1 wd/2 pgs	1 wd/5 pgs	<1wd/5pg	أقل من كلمة في كل 5 أوجه أكثر من 5 كلمات في كل وجه
12. HOW ENJOYABLE was the material to read?	ضع دائرة حول تقييمك لدرجة متعة قراءة هذه المادة:						
	Very enjoyable	5	4	3	2	1	Very unenjoyable
	ممتعة جداً						مملة جداً
13. HOW INTERESTING was the reading material?	ضع دائرة حول تقييمك لدرجة اهتمام قراءة هذه المادة:						
	Very interesting	5	4	3	2	1	Very boring
	مبهتة جداً						غير مبهت أبداً
14. HOW ESSENTIAL was it to read the material?	ضع دائرة حول تقييمك لدرجة ضرورة قراءة هذه المادة، في الحالة التي كنت فيها:						
	Really Essential	5	4	3	2	1	Unessential
	ضرورية						بدون أية ضرورة

1. Write the DAY when the reading took place:	يوم القراءة:						
2. Write the DATE when the reading took place:	تاريخ القراءة:						
3. Specify the LANGUAGE of the reading material:	ENGLISH	OR	أو	اللغة العربية	حدد لغة المادة المقروءة:		
4. Describe the TYPE of reading material:	حدد نوع المادة المقروءة:						
5. Write the TITLE or brief description:	اكتب عنوان المادة المقروءة أو وصفها الموجز:						
6. WHEN did you read the material?	ضع دائرة حول أوقات قراءة هذه المادة:						
	12am-6am	6am-8am	8am-12pm	12pm-4pm	4pm-9pm	9pm-12am	ليلاً صباحاً
7. HOW MUCH TIME in total did you spent reading the material?	ضع دائرة حول مدة قراءة هذه المادة الإجمالية:						
	Under 10min	10-20min	20-40min	40min-1hour	1-2hours	Over 2hours	أكثر من ساعتين أقل من 10 دقيقة
8. WHY did you read the material?	حدد أسباب قراءة هذه المادة:						
9. WHERE did you read the material?	حدد مكان قراءة هذه المادة:						
10. HOW EASY was the material to read?	ضع دائرة حول تقييمك لسهولة قراءة هذه المادة:						
	Very easy	5	4	3	2	1	Very hard
	سهلة جداً						صعبة جداً
11. HOW OFTEN did you use the DICTIONARY or glossary?	في استخدام القاموس خلال القراءة						
	>5wds/pg	2-5 wds/pg	1 wd/pg	1 wd/2 pgs	1 wd/5 pgs	<1wd/5pg	أقل من كلمة في كل 5 أوجه أكثر من 5 كلمات في كل وجه
12. HOW ENJOYABLE was the material to read?	ضع دائرة حول تقييمك لدرجة متعة قراءة هذه المادة:						
	Very enjoyable	5	4	3	2	1	Very unenjoyable
	ممتعة جداً						مملة جداً
13. HOW INTERESTING was the reading material?	ضع دائرة حول تقييمك لدرجة اهتمام قراءة هذه المادة:						
	Very interesting	5	4	3	2	1	Very boring
	مبهتة جداً						غير مبهت أبداً
14. HOW ESSENTIAL was it to read the material?	ضع دائرة حول تقييمك لدرجة ضرورة قراءة هذه المادة، في الحالة التي كنت فيها:						
	Really Essential	5	4	3	2	1	Unessential
	ضرورية						بدون أية ضرورة

1. Write the DAY when the reading took place:	يوم القراءة:						
2. Write the DATE when the reading took place:	تاريخ القراءة:						
3. Specify the LANGUAGE of the reading material:	ENGLISH	OR	أو	اللغة العربية	حدد لغة المادة المقروءة:		
4. Describe the TYPE of reading material:	حدد نوع المادة المقروءة:						
5. Write the TITLE or brief description:	اكتب عنوان المادة المقروءة أو وصفها الموجز:						
6. WHEN did you read the material?	ضع دائرة حول أوقات قراءة هذه المادة:						
	12am-6am	6am-8am	8am-12pm	12pm-4pm	4pm-9pm	9pm-12am	ليلاً صباحاً
7. HOW MUCH TIME in total did you spent reading the material?	ضع دائرة حول مدة قراءة هذه المادة الإجمالية:						
	Under 10min	10-20min	20-40min	40min-1hour	1-2hours	Over 2hours	أكثر من ساعتين أقل من 10 دقيقة
8. WHY did you read the material?	حدد أسباب قراءة هذه المادة:						
9. WHERE did you read the material?	حدد مكان قراءة هذه المادة:						
10. HOW EASY was the material to read?	ضع دائرة حول تقييمك لسهولة قراءة هذه المادة:						
	Very easy	5	4	3	2	1	Very hard
	سهلة جداً						صعبة جداً
11. HOW OFTEN did you use the DICTIONARY or glossary?	في استخدام القاموس خلال القراءة						
	>5wds/pg	2-5 wds/pg	1 wd/pg	1 wd/2 pgs	1 wd/5 pgs	<1wd/5pg	أقل من كلمة في كل 5 أوجه أكثر من 5 كلمات في كل وجه
12. HOW ENJOYABLE was the material to read?	ضع دائرة حول تقييمك لدرجة متعة قراءة هذه المادة:						
	Very enjoyable	5	4	3	2	1	Very unenjoyable
	ممتعة جداً						مملة جداً
13. HOW INTERESTING was the reading material?	ضع دائرة حول تقييمك لدرجة اهتمام قراءة هذه المادة:						
	Very interesting	5	4	3	2	1	Very boring
	مبهتة جداً						غير مبهت أبداً
14. HOW ESSENTIAL was it to read the material?	ضع دائرة حول تقييمك لدرجة ضرورة قراءة هذه المادة، في الحالة التي كنت فيها:						
	Really Essential	5	4	3	2	1	Unessential
	ضرورية						بدون أية ضرورة

Appendix-12: Weekly Reading Diary Sheet - Stage-3 Main Study [see Section-3.6.1]

1. Write the WEEK when the reading took place:								
2. Specify the LANGUAGE of the reading material: ENGLISH OR اللغة العربية								
3. Describe the TYPE of reading material:								
College/Work Book الدراسة أو النشاط	Teacher's Lecture ملخص المدرس للحصة	Student's Own Notes ملخص الطالب للحصة	Newspaper جريدة	Magazine مجلة	Book with Translation كتاب مع ترجمة	Graded Reader قصة من سلسلة تعليمية مترجمة	Simplified Book with Pictures كتاب مبسّط اللغة فيه صور كثيرة	Simplified Book with FEW Pictures كتاب مبسّط اللغة فيه صور قليلة
Internet الإنترنت	Computer/ Mobile Offline كمبيوتر أو جوال بدون الإنترنت	Film with Subtitles - circle the appropriate options:	English Film or أفلام عربية	with English Subtitles or أو مع ترجمة عربية	فيلم أو برنامج مع ترجمة - ضع دائرة حول الخيارات المناسبة:	Full Novel رواية أو قصة بلغة عادية غير مبسّطة	Specify another reading material	حدد نوعاً آخر من المادة المقروءة
4. Write the TITLE or brief description:								
5. HOW MUCH TIME did you spent reading the material DURING THE WEEK?								
أكثر من خمس ساعات 4-5hours Over 5hours 3-4hours 2-3hours 1-2hours 30min-1hour Under 30min أقل من 30 دقيقة								
6. HOW OFTEN did you read the material DURING THE WEEK?								
أكثر من سبع جلسات 7 more than 7 6 5 4 3 2 1								
7. HOW EASY was the material to read?								
صعبة جداً 1 Very hard 2 3 4 5 سهلة جداً 5 Very easy								
8. HOW OFTEN did you use the DICTIONARY? ELECTRONIC or BOOK?								
أقل من كلمة في كل 5 <1wd/5pg 1 wd/5 pgs 1 wd/2 pgs 1 wd/pg 2-5 wds/pg أكثر من 5 كلمات في كل وجه >5wds/pg								
9. HOW ENJOYABLE was the material to read?								
ممتعة جداً 5 4 3 2 1 Very unenjoyable صعبة جداً 1								
10. HOW INTERESTING was the reading material?								
غير ممتع أبداً 1 Very boring 2 3 4 5 ممتع جداً 5 Very interesting								
11. HOW ESSENTIAL was it to read the material?								
بدون أية ضرورة 1 Unessential 2 3 4 5 ضرورية جداً 5 Really Essential								

1. Write the WEEK when the reading took place:								
2. Specify the LANGUAGE of the reading material: ENGLISH OR اللغة العربية								
3. Describe the TYPE of reading material:								
College/Work Book الدراسة أو النشاط	Teacher's Lecture ملخص المدرس للحصة	Student's Own Notes ملخص الطالب للحصة	Newspaper جريدة	Magazine مجلة	Book with Translation كتاب مع ترجمة	Graded Reader قصة من سلسلة تعليمية مترجمة	Simplified Book with Pictures كتاب مبسّط اللغة فيه صور كثيرة	Simplified Book with FEW Pictures كتاب مبسّط اللغة فيه صور قليلة
Internet الإنترنت	Computer/ Mobile Offline كمبيوتر أو جوال بدون الإنترنت	Film with Subtitles - circle the appropriate options:	English Film or أفلام عربية	with English Subtitles or أو مع ترجمة عربية	فيلم أو برنامج مع ترجمة - ضع دائرة حول الخيارات المناسبة:	Full Novel رواية أو قصة بلغة عادية غير مبسّطة	Specify another reading material	حدد نوعاً آخر من المادة المقروءة
4. Write the TITLE or brief description:								
5. HOW MUCH TIME did you spent reading the material DURING THE WEEK?								
أكثر من خمس ساعات 4-5hours Over 5hours 3-4hours 2-3hours 1-2hours 30min-1hour Under 30min أقل من 30 دقيقة								
6. HOW OFTEN did you read the material DURING THE WEEK?								
أكثر من سبع جلسات 7 more than 7 6 5 4 3 2 1								
7. HOW EASY was the material to read?								
صعبة جداً 1 Very hard 2 3 4 5 سهلة جداً 5 Very easy								
8. HOW OFTEN did you use the DICTIONARY? ELECTRONIC or BOOK?								
أقل من كلمة في كل 5 <1wd/5pg 1 wd/5 pgs 1 wd/2 pgs 1 wd/pg 2-5 wds/pg أكثر من 5 كلمات في كل وجه >5wds/pg								
9. HOW ENJOYABLE was the material to read?								
ممتعة جداً 5 4 3 2 1 Very unenjoyable صعبة جداً 1								
10. HOW INTERESTING was the reading material?								
غير ممتع أبداً 1 Very boring 2 3 4 5 ممتع جداً 5 Very interesting								
11. HOW ESSENTIAL was it to read the material?								
بدون أية ضرورة 1 Unessential 2 3 4 5 ضرورية جداً 5 Really Essential								

Appendix-13: Survey of Exposure to English Learning - Stage-3 Main Study [see Section-3.7.1]

Age:

Please write where you have studied English (including any private tuition), but NOT AT THIS COLLEGE:

أين درست الانكليزية (مدرسة أو كلية أو معهد أو غيرها) ما حدا هذه الكلية؟

Name of Institution	Type of Institution	Country	Year Joined	Months or Years spent there	How much of the teaching was in English? (%)	Hours of English lessons every week	Months of study of English every year	How much English outside the English class? (%)
اسم المركز التعليمي	نوع المركز التعليمي والمرحلة	البلاد	تاريخ الدخول (أو الصف)	مدة الدراسة (كم سنة أو شهرا)	نسبة التعليم بالانكليزية	عدد ساعات دروس اللغة الانكليزية في الأسبوع	عدد أشهر دراسة اللغة الانكليزية في السنة	نسبة استخدام الانكليزية خارج الدروس الانكليزية

Please write about any international English exams you have taken.

(امتحانات اللغة الانكليزية الدولية)

Name of Exam	Year	Result	Maximum Possible Score
اسم الامتحانات	العام	النتيجة أو النسبة	أعلى نسبة ممكنة في الامتحان

Appendix-13: Survey of Exposure to English Learning - Stage-3 Main Study [see Section-3.7.1] (cont.)

4 Please write about the places you have travelled to, during which you used English:

(إلى أين سافرت واستخدمت الإنجليزية خلال هذا السفر؟)

Country البلد	Year of Trip عام الرحلة	Duration of Trip مدة الرحلة	Purpose السبب أو الهدف	% Time using نسبة استخدام اللغة الإنجليزية أثناء السفر

5 On average, how OFTEN do you watch English films or programmes?

عدد جلسات مشاهدة برامج إنجليزية أو أفلام إنجليزية

More than once a day	Once a day	More than once a week	Once a week	More than once every 2 weeks	Once every 2 weeks	More than once a month	Once a month	Less than once a month
أكثر من مرة في اليوم	مرة واحدة في اليوم	أكثر من مرة في الأسبوع	مرة واحدة في الأسبوع	أكثر من مرة في الأسبوعين	مرة واحدة في الأسبوعين	أكثر من مرة في الشهر	مرة واحدة في الشهر	أقل من مرة واحدة في الشهر

6 On average, how much TIME do you spend watching English films or programmes, EVERY WEEK?

مدة مشاهدة برامج إنجليزية أو أفلام إنجليزية في الأسبوع

أقل من 30 دقيقة في الأسبوع	Less than 30 min a week	30min to 1 hour	1-2 hours	2-3 hours	3-5 hours	5-10 hours	10-20 hours	More than 20 hours a week	أكثر من 20 ساعة في الأسبوع
----------------------------	-------------------------	-----------------	-----------	-----------	-----------	------------	-------------	---------------------------	----------------------------

7 On average, how OFTEN do you listen to English radio or music?

عدد جلسات استماع لـصوتيات إنجليزية

More than once a day	Once a day	More than once a week	Once a week	More than once every 2 weeks	Once every 2 weeks	More than once a month	Once a month	Less than once a month
أكثر من مرة في اليوم	مرة واحدة في اليوم	أكثر من مرة في الأسبوع	مرة واحدة في الأسبوع	أكثر من مرة في الأسبوعين	مرة واحدة في الأسبوعين	أكثر من مرة في الشهر	مرة واحدة في الشهر	أقل من مرة واحدة في الشهر

8 On average, how much TIME do you spend listening to English radio or music, EVERY WEEK?

مدة استماع لـصوتيات إنجليزية في الأسبوع

أقل من 30 دقيقة في الأسبوع	Less than 30 min a week	30min to 1 hour	1-2 hours	2-3 hours	3-5 hours	5-10 hours	10-20 hours	More than 20 hours a week	أكثر من 20 ساعة في الأسبوع
----------------------------	-------------------------	-----------------	-----------	-----------	-----------	------------	-------------	---------------------------	----------------------------

9 How much English do you use at work?

استخدام الإنجليزية في أمكنة العمل

Name of Company اسم الشركة	Country البلد	Year Joined سنة بداية العمل	Months or Years spent there مدة العمل	Hours of Work every week ساعات العمل في الأسبوع	How much English was used during the Work? (%) نسبة استخدام الإنجليزية

Appendix-14: Results of English Placement Test [see Section-4.1]

ENGLISH PLACEMENT TESTS

DESCRIPTIVE STATISTICS:	STAGE-2 MAIN PILOT		STAGE-3 MAIN STUDY		
	Main Pilot Cohort	All Students in Batch	Experimental Cohort	Control Cohort	All Students in Batch
Number of Participants in the Study	52	52	49	36	85
Number who took the Placement Test	48	593	45	32	748
Mean (%)	42	43	43	37	37
Standard Deviation (%)	15	14	15	18	15
Minimum (%)	12	15	10	3	0
Lower Quartile Q1 (%)	33	34	32	25	28
Median Q2 (%)	41	40	45	33	35
Upper Quartile Q3 (%)	52	50	54	50	45
Maximum (%)	88	92	85	85	95
Range (%)	76	77	75	82	95

COMPARING GROUPS:	p (<0.05 for significant* difference in results)	U* value	Z* value	r Effect Size**
Main Pilot Cohort & All Students in Batch	0.818	13948	-0.23	0.01
Main Pilot Cohort & Experimental Cohort	0.658	1023	-0.44	0.05
Main Pilot Cohort & Control Cohort	0.093	597	-1.68	0.19
Experimental Cohort & Control Cohort	0.077	549	-1.77	0.20
Experimental Cohort & All Students in Batch	0.004	292630	-2.90	0.11
Control Cohort & All Students in Batch	0.803	11657	-0.25	0.01

*Mann-Whitney U Test for non-normally distributed independent samples. See Pallant (2010:227-230).

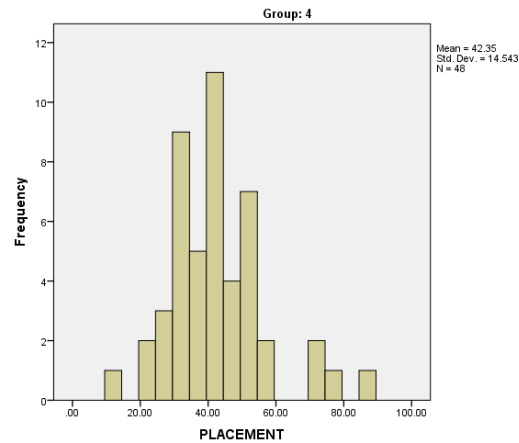
**Effect Size $r = |Z| / \sqrt{N_{total}}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:230).

Note: It can be seen above that a few participants did not take the placement test, simply because they started classes a day later. These were subsequently allocated to preserve similar class sizes.

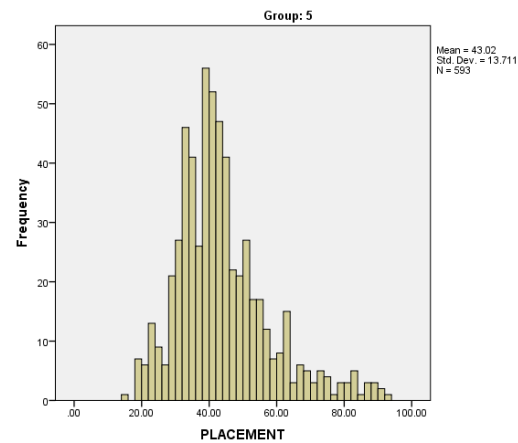
Appendix-14: Results of English Placement Test [see Section-4.1] (cont.)

STAGE-2 MAIN PILOT

Main Pilot Cohort

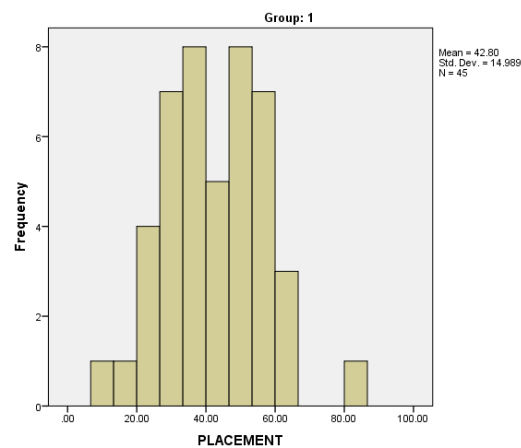


All Students in Batch

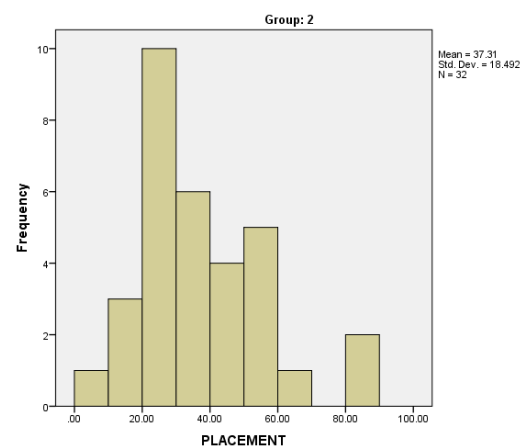


STAGE-3 MAIN STUDY

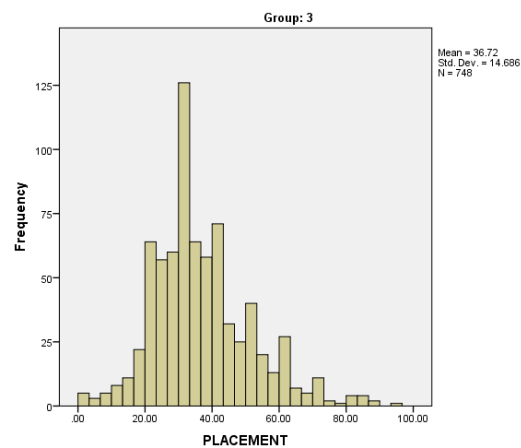
Experimental Cohort



Control Cohort



All Students in Batch



Appendix-15: Results of the Stage-2 Main Pilot Vocabulary Tests [see Section-3.4.7]

TEST-1 Based on NATION (2009)
SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.805	10.0	0.777	11.0	0.077	-1.77	0.17	1.0	10%
	Correctness x Certainty	0.852	8.2	0.846	9.2	0.008	-2.67	0.26	1.0	12%
	Certainty only	0.923	13.5	0.919	14.0	0.060	-1.88	0.18	0.5	4%
6000-10000wd 25 items	Correctness only	0.749	5.0	0.798	5.0	0.325	-0.99	0.10	0.0	0%
	Correctness x Certainty	0.733	2.8	0.869	3.0	0.341	-0.95	0.09	0.2	6%
	Certainty only	0.927	5.8	0.934	6.8	0.188	-1.32	0.13	1.0	17%

TEST-2 Based on NATION (2009)
TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.821	6.0	0.853	6.5	0.000	-3.69	0.36	0.5	8%
	Correctness x Certainty	0.846	5.0	0.872	6.3	0.000	-3.65	0.36	1.3	27%
	Certainty only	0.870	8.0	0.889	10.0	0.006	-2.76	0.27	2.0	25%
6000-10000wd 25 items	Correctness only	0.426	0.0	0.899	0.0	0.019	-2.35	0.23	0.0	Pretest = 0.0
	Correctness x Certainty	0.487	0.0	0.924	0.0	0.005	-2.83	0.28	0.0	Pretest = 0.0
	Certainty only	0.829	1.3	0.905	2.2	0.013	-2.48	0.24	0.8	63%

TEST-3 Based on NATION (2009)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.806	8.5	0.815	9.0	0.011	-2.55	0.25	0.5	6%
	Correctness x Certainty	0.858	6.3	0.862	7.3	0.000	-3.88	0.38	1.0	16%
	Certainty only	0.931	12.2	0.933	14.0	0.002	-3.08	0.30	1.8	15%
6000-10000wd 25 items	Correctness only	0.712	3.0	0.802	3.0	0.687	-0.40	0.04	0.0	0%
	Correctness x Certainty	0.744	1.7	0.893	1.8	0.207	-1.26	0.12	0.2	10%
	Certainty only	0.906	6.7	0.947	5.3	0.687	-0.40	0.04	-1.3	-20%

TEST-4 Based on SCHMITT ET AL (2001)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.913	7.5	0.918	8.0	0.009	-2.61	0.26	0.5	7%
	Correctness x Certainty	0.939	5.2	0.940	6.3	0.001	-3.34	0.33	1.2	23%
	Certainty only	0.961	14.2	0.956	15.0	0.067	-1.83	0.18	0.8	6%
3000wd 30 items	Correctness only	0.868	4.0	0.890	5.0	0.146	-1.45	0.14	1.0	25%
	Correctness x Certainty	0.892	3.0	0.924	3.3	0.029	-2.18	0.21	0.3	11%
	Certainty only	0.945	9.3	0.944	9.0	0.352	-0.93	0.09	-0.3	-4%
Academic wd 30 items	Correctness only	0.901	4.0	0.911	6.0	0.003	-3.01	0.30	2.0	50%
	Correctness x Certainty	0.923	2.8	0.933	4.3	0.001	-3.41	0.33	1.5	53%
	Certainty only	0.945	9.2	0.956	8.8	0.042	-2.03	0.20	-0.3	-4%
5000wd 30 items	Correctness only	0.799	2.0	0.813	3.0	0.052	-1.94	0.19	1.0	50%
	Correctness x Certainty	0.808	1.0	0.871	1.7	0.026	-2.23	0.22	0.7	67%
	Certainty only	0.936	5.0	0.947	4.7	0.231	-1.20	0.12	-0.3	-7%
10000wd 30 items	Correctness only	0.715	0.0	0.717	1.0	0.597	-0.53	0.05	1.0	Pretest = 0.0
	Correctness x Certainty	0.741	0.0	0.789	0.3	0.799	-0.26	0.03	0.3	Pretest = 0.0
	Certainty only	0.937	2.3	0.954	2.0	0.646	-0.46	0.05	-0.3	-14%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill; Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill; Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{(N_{pretest} + N_{posttest})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = ((posttest score – pretest score) / pretest score) x 100. Note if pretest score = 0.0, then Percentage Change cannot be calculated.

Appendix-15: Results of the Stage-2 Main Pilot Vocabulary Tests [see Section3.4.7]

(cont.)

TEST-1 SPELLING

Based on NATION (2009)

MAIN PILOT n= 52

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	10.3	4.5	8.0	10.0	13.8	11.2	4.3	8.0	11.0	14.0
	Correctness x Certainty	8.3	4.1	5.2	8.2	11.6	9.3	4.2	5.7	9.2	12.6
	Certainty only	12.6	5.2	7.7	13.5	16.6	13.6	5.0	10.0	14.0	17.3
6000-10000wd 25 items	Correctness only	5.0	3.6	3.0	5.0	6.0	5.4	4.0	2.0	5.0	8.0
	Correctness x Certainty	3.1	2.4	1.1	2.8	4.2	3.5	3.4	1.0	3.0	4.9
	Certainty only	7.0	5.0	3.3	5.8	10.6	7.9	5.3	3.7	6.8	10.8

TEST-2 TRANSLATION

Based on NATION (2009)

MAIN PILOT n= 52

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	6.4	3.6	4.0	6.0	8.0	7.5	4.2	4.0	6.5	10.0
	Correctness x Certainty	5.9	3.5	3.1	5.0	7.9	6.9	4.1	3.5	6.3	9.0
	Certainty only	8.7	4.3	5.7	8.0	10.7	9.8	4.8	5.8	10.0	12.7
6000-10000wd 25 items	Correctness only	0.6	1.0	0.0	0.0	1.0	1.2	2.8	0.0	0.0	1.0
	Correctness x Certainty	0.5	0.8	0.0	0.0	0.7	1.0	2.8	0.0	0.0	1.0
	Certainty only	2.2	2.6	0.3	1.3	3.0	3.1	4.0	1.0	2.2	3.7

TEST-3 MEANING

Based on NATION (2009)

MAIN PILOT n= 52

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	8.2	4.4	5.0	8.5	12.0	9.3	4.6	5.0	9.0	12.8
	Correctness x Certainty	6.5	4.1	3.3	6.3	9.8	7.8	4.4	4.3	7.3	10.7
	Certainty only	11.7	5.1	7.3	12.2	16.2	13.2	5.2	9.1	14.0	17.0
6000-10000wd 25 items	Correctness only	3.5	3.0	1.0	3.0	5.8	3.8	3.7	1.0	3.0	5.0
	Correctness x Certainty	1.9	1.9	0.3	1.7	2.9	2.4	3.3	0.7	1.8	3.3
	Certainty only	6.4	4.0	2.8	6.7	9.6	6.6	5.2	2.3	5.3	9.7

TEST-4 MEANING

Based on SCHMITT ET AL (2001)

MAIN PILOT n= 52

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
2000wd level 30 items	Correctness only	9.2	7.0	4.0	7.5	12.0	10.4	7.2	5.0	8.0	15.8
	Correctness x Certainty	7.2	6.5	3.0	5.2	10.2	8.5	6.8	3.4	6.3	12.4
	Certainty only	13.6	7.6	7.0	14.2	19.3	14.6	7.3	9.0	15.0	20.4
3000wd level 30 items	Correctness only	5.8	5.1	2.0	4.0	10.0	6.4	5.6	2.3	5.0	8.0
	Correctness x Certainty	4.3	4.4	1.0	3.0	6.3	4.9	5.2	1.7	3.3	6.0
	Certainty only	9.4	6.4	3.0	9.3	15.6	9.7	6.5	3.8	9.0	14.2
Academic wd level 30 items	Correctness only	6.4	5.9	2.0	4.0	10.0	7.8	6.4	3.0	6.0	12.0
	Correctness x Certainty	5.1	5.3	1.1	2.8	7.3	6.3	6.0	1.7	4.3	9.9
	Certainty only	9.5	6.4	3.8	9.2	14.2	10.8	7.3	4.1	8.8	17.2
5000wd level 30 items	Correctness only	2.6	3.0	0.0	2.0	4.0	3.4	3.6	1.0	3.0	5.0
	Correctness x Certainty	1.7	2.1	0.0	1.0	2.2	2.2	2.9	0.3	1.7	3.0
	Certainty only	6.0	5.0	2.1	5.0	7.8	6.5	5.7	2.0	4.7	10.9
10000wd level 30 items	Correctness only	1.3	2.0	0.0	0.0	2.0	1.5	2.0	0.0	1.0	3.0
	Correctness x Certainty	0.6	1.0	0.0	0.0	0.7	0.8	1.4	0.0	0.3	1.3
	Certainty only	3.6	3.8	0.7	2.3	5.3	4.0	4.7	0.7	2.0	6.8

- Nation, I.S.P. (2009). "Vocabulary size test." Available from: <http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

Appendix-16: Comparing Quantitative Data of the Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49) [see Section-3.9.1]

			INTERVIEW PARTICIPANTS						
			n= 12						
			Mean	Standard Deviation	Minimum	Lower Quartile Q1	Median Q2	Upper Quartile Q3	Maximum Range
AGE (years)			23	3.7	20	21	22	25	34
END OF SEMESTER EXAM (%)			85	7.8	74	77	88	91	98
TEST-1 SPELLING Based on NATION (2009) (score/25)	1000-5000wd	Pretest	11.9	3.6	7.0	9.3	11.0	14.3	20.0
	<i>Correctness only</i>	Posttest	13.1	3.3	7.0	11.0	13.0	16.8	17.0
	1000-5000wd	Pretest	9.7	2.7	5.0	7.8	10.3	11.5	13.7
	<i>Correctness × Certainty</i>	Posttest	10.8	2.7	6.3	9.0	10.0	13.9	14.7
	1000-5000wd	Pretest	13.6	3.4	9.3	10.3	14.2	15.9	20.3
	<i>Certainty Only</i>	Posttest	14.8	2.7	9.7	12.4	15.2	17.4	18.3
TEST-2 TRANSLATION (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	8.3	2.3	3.0	7.0	8.5	9.8	12.0
	<i>Correctness only</i>	Posttest	9.9	2.8	3.0	8.0	10.5	12.0	13.0
	1000-5000wd	Pretest	7.9	2.1	3.0	6.8	8.0	9.0	11.0
	<i>Correctness × Certainty</i>	Posttest	9.2	2.7	3.0	7.4	9.7	11.6	12.0
	1000-5000wd	Pretest	10.9	2.6	5.3	9.3	11.5	12.8	15.0
	<i>Certainty Only</i>	Posttest	11.8	3.3	5.3	9.4	12.5	13.7	17.0
TEST-3 MEANING (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	11.5	3.3	7.0	9.3	11.0	13.8	18.0
	<i>Correctness only</i>	Posttest	11.3	3.8	4.0	8.3	12.0	14.8	16.0
	1000-5000wd	Pretest	9.5	3.0	4.3	7.5	8.5	12.0	14.7
	<i>Correctness × Certainty</i>	Posttest	9.9	3.6	3.3	7.0	10.4	13.5	14.0
	1000-5000wd	Pretest	14.1	2.6	8.3	12.5	15.0	15.9	17.7
	<i>Certainty Only</i>	Posttest	15.2	2.7	9.3	13.3	15.7	16.8	19.3
TEST-4 MEANING (/30) Based on SCHMITT ET AL (2001) (score/30)	2000wd level	Pretest	9.8	5.4	3.0	5.0	8.5	14.5	20.0
	<i>Correctness only</i>	Posttest	13.7	5.4	4.0	9.8	13.0	17.0	23.0
	2000wd level	Pretest	7.2	3.9	2.3	3.7	5.9	10.9	14.3
	<i>Correctness × Certainty</i>	Posttest	11.2	4.7	3.3	7.5	11.4	14.8	19.3
	2000wd level	Pretest	14.2	4.6	7.0	11.5	14.3	17.1	23.7
	<i>Certainty Only</i>	Posttest	17.1	4.7	8.3	14.5	17.2	21.0	24.3
	3000wd level	Pretest	7.6	4.1	2.0	4.3	7.0	11.3	15.0
	<i>Correctness only</i>	Posttest	8.3	4.5	2.0	4.0	10.0	11.0	15.0
	3000wd level	Pretest	5.8	3.6	0.7	2.9	5.4	8.2	12.7
	<i>Correctness × Certainty</i>	Posttest	6.7	4.1	1.3	2.6	6.5	10.0	13.0
	3000wd level	Pretest	10.2	5.0	4.0	5.3	10.0	13.2	21.3
	<i>Certainty Only</i>	Posttest	11.7	5.6	3.0	7.3	12.0	15.5	22.7
	Academic wd level	Pretest	7.8	5.6	1.0	3.3	6.5	12.0	18.0
	<i>Correctness only</i>	Posttest	9.3	5.4	0.0	5.3	8.0	14.0	19.0
	Academic wd level	Pretest	5.5	4.7	0.3	2.2	3.2	10.3	14.0
	<i>Correctness × Certainty</i>	Posttest	7.1	5.5	0.0	3.3	4.7	13.5	16.7
	Academic wd level	Pretest	10.5	6.0	1.3	5.8	10.2	15.9	19.7
	<i>Certainty Only</i>	Posttest	12.9	6.6	1.3	7.8	11.8	16.5	25.3

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:
<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].
 - Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

Appendix-16: Comparing Quantitative Data of the Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49) [see Section-3.9.1] (cont.)

			ALL PARTICIPANTS OF EXPERIMENTAL COHORT n= 49						
			Mean	Standard Deviation	Minimum	Lower Quartile Q1	Median Q2	Upper Quartile Q3	Maximum Range
AGE (years)			23	3.2	20	22	22	23	36
END OF SEMESTER EXAM (%)			80	12.3	51	70	83	90	99
TEST-1 SPELLING Based on NATION (2009) (score/25)	1000-5000wd	Pretest	10.4	3.6	3.0	8.5	10.0	12.0	20.0
	Correctness only	Posttest	11.8	3.4	7.0	9.0	11.0	14.0	21.0
	1000-5000wd	Pretest	8.0	3.2	2.3	5.5	7.7	10.5	14.3
	Correctness × Certainty	Posttest	9.8	3.4	4.3	7.3	9.3	12.0	20.0
	1000-5000wd	Pretest	12.0	4.1	4.0	8.7	12.3	15.8	20.3
	Certainty Only	Posttest	14.3	3.7	6.7	11.2	14.3	17.0	22.7
TEST-2 TRANSLATION (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	6.8	2.8	2.0	4.0	7.0	9.0	13.0
	Correctness only	Posttest	8.7	3.1	2.0	7.0	9.0	11.0	16.0
	1000-5000wd	Pretest	6.3	2.7	2.0	4.0	6.0	8.2	13.0
	Correctness × Certainty	Posttest	8.0	3.0	2.0	5.5	8.3	10.5	15.0
	1000-5000wd	Pretest	9.0	3.2	4.0	6.0	8.7	11.7	16.3
	Certainty Only	Posttest	11.0	3.4	4.3	8.3	10.7	13.3	17.7
TEST-3 MEANING (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	9.4	3.7	2.0	7.0	9.0	12.0	18.0
	Correctness only	Posttest	10.6	3.5	4.0	8.0	10.0	13.0	20.0
	1000-5000wd	Pretest	7.7	3.3	1.0	5.5	7.3	9.8	15.3
	Correctness × Certainty	Posttest	8.8	3.3	3.0	6.0	8.3	11.0	18.3
	1000-5000wd	Pretest	12.5	3.6	5.0	9.3	13.0	15.7	18.0
	Certainty Only	Posttest	13.9	4.0	6.0	11.3	14.3	16.7	20.0
TEST-4 MEANING (/30) Based on SCHMITT ET AL (2001) (score/30)	2000wd level	Pretest	9.0	5.7	1.0	5.0	7.0	12.5	29.0
	Correctness only	Posttest	11.3	6.2	1.0	6.5	11.0	14.5	29.0
	2000wd level	Pretest	6.6	5.1	0.7	3.2	5.0	9.0	29.0
	Correctness × Certainty	Posttest	8.9	6.0	1.0	4.5	7.0	11.8	29.0
	2000wd level	Pretest	13.0	5.4	1.7	9.5	13.3	16.2	30.0
	Certainty Only	Posttest	15.6	5.7	1.7	11.3	15.7	20.2	30.0
	3000wd level	Pretest	6.5	4.6	1.0	3.0	5.0	9.0	24.0
	Correctness only	Posttest	7.9	5.6	1.0	4.0	7.0	11.0	29.0
	3000wd level	Pretest	4.6	4.1	0.3	1.7	3.7	6.5	22.3
	Correctness × Certainty	Posttest	5.9	4.9	0.7	2.3	4.3	8.5	27.7
	3000wd level	Pretest	9.7	5.5	0.7	5.3	8.7	13.2	26.0
	Certainty Only	Posttest	11.4	5.8	0.7	8.3	10.7	15.2	28.7
	Academic wd level	Pretest	7.1	5.3	1.0	3.0	6.0	9.0	25.0
	Correctness only	Posttest	6.3	6.0	0.0	2.0	5.0	9.5	27.0
	Academic wd level	Pretest	5.1	4.7	0.3	2.0	3.0	7.2	23.7
	Correctness × Certainty	Posttest	6.4	5.6	0.0	2.2	4.3	8.3	26.0
	Academic wd level	Pretest	9.9	5.7	0.7	5.3	8.7	14.8	24.0
	Certainty Only	Posttest	12.6	6.7	1.0	7.8	12.0	17.5	28.0

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:
<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].
- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

Appendix-16: Comparing Quantitative Data of the Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49) [see Section-3.9.1] (cont.)

Comparing Data Profile of Interview Participants (n=12) with all Participants of the Experimental Cohort (n=49)			COMPARING GROUPS:			
			p (<0.05 for significant* difference in results)	U* value	Z* value	r Effect Size**
AGE (years)			0.867	285	-0.17	0.02
END OF SEMESTER EXAM (%)			0.224	227	-1.22	0.16
TEST-1 SPELLING Based on NATION (2009) (score/25)	1000-5000wd	Pretest	0.208	225	-1.26	0.16
	Correctness only	Posttest	0.169	219	-1.38	0.18
	1000-5000wd	Pretest	0.081	198	-1.74	0.22
	Correctness × Certainty	Posttest	0.134	212	-1.50	0.19
	1000-5000wd	Pretest	0.318	239	-1.00	0.13
	Certainty Only	Posttest	0.543	261	-0.61	0.08
TEST-2 TRANSLATION (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	0.065	193	-1.85	0.24
	Correctness only	Posttest	0.116	208	-1.57	0.20
	1000-5000wd	Pretest	0.048	185	-1.98	0.25
	Correctness × Certainty	Posttest	0.154	216	-1.43	0.18
	1000-5000wd	Pretest	0.058	190	-1.90	0.24
	Certainty Only	Posttest	0.364	244	-0.91	0.12
TEST-3 MEANING (/25) Based on NATION (2009) (score/25)	1000-5000wd	Pretest	0.074	196	-1.79	0.23
	Correctness only	Posttest	0.397	248	-0.85	0.11
	1000-5000wd	Pretest	0.112	207	-1.59	0.20
	Correctness × Certainty	Posttest	0.234	229	-1.19	0.15
	1000-5000wd	Pretest	0.170	219	-1.37	0.18
	Certainty Only	Posttest	0.322	240	-0.99	0.13
TEST-4 MEANING (/30) Based on SCHMITT ET AL (2001) (score/30)	2000wd level	Pretest	0.590	265	-0.54	0.07
	Correctness only	Posttest	0.136	212	-1.49	0.19
	2000wd level	Pretest	0.403	248	-0.84	0.11
	Correctness × Certainty	Posttest	0.086	200	-1.71	0.22
	2000wd level	Pretest	0.419	250	-0.81	0.10
	Certainty Only	Posttest	0.364	244	-0.91	0.12
	3000wd level	Pretest	0.271	234	-1.10	0.14
	Correctness only	Posttest	0.585	264	-0.55	0.07
	3000wd level	Pretest	0.154	216	-1.43	0.18
	Correctness × Certainty	Posttest	0.484	256	-0.70	0.09
	3000wd level	Pretest	0.611	266	-0.51	0.07
	Certainty Only	Posttest	0.806	281	-0.25	0.03
	Academic wd level	Pretest	0.655	270	-0.45	0.06
	Correctness only	Posttest	0.056	189	-1.91	0.24
	Academic wd level	Pretest	0.624	267	-0.49	0.06
	Correctness × Certainty	Posttest	0.650	269	-0.45	0.06
	Academic wd level	Pretest	0.650	269	-0.45	0.06
	Certainty Only	Posttest	0.849	284	-0.19	0.02

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

*Mann-Whitney U Test for non-normally distributed independent samples. See Pallant (2010:227-230).

**Effect Size $r = |Z| / \sqrt{N_{total}}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:230).

Appendix-17: In-Depth Interview Schedule [see Section-3.9.2.1]

BRIEFING

Once again, I thank you very much for agreeing to be interviewed.

As you know, your information will be extremely useful to my research and my future teaching.

Also, as I mentioned earlier, I would like to record the interview.

INTRODUCTION & INITIAL QUESTIONS

How are your current studies?

What are you studying now in English?

How is college life, now you are no longer new to the college?

Now, I ask you to please be honest and frank. Please tell me what you really think.

You don't have to say nice things just to make me happy.

I'm no longer teaching you and so I don't have anything to do with your college grades.

Also be as detailed as you can. Tell me everything.

We chose to meet now, because you said you were free, so please relax and take it easy.

Don't feel rushed or pressured, because we have time.

Please remember that without your honest and complete feedback, there will be no benefit to my research into reading and vocabulary, and there will be no benefit to my future teaching of these.

CONTENT QUESTION #1

WHAT DID YOU THINK ABOUT THE READING PROGRAMME?

CORE PROBES

In what ways was the Reading Programme useful?

What did you think about the in-class silent reading sessions?

What did you think about filling in the Reading Diary sheets?

Describe your feelings about reading during the 12 weeks.

POSSIBLE FOLLOW-UP PROBES

In what ways was the Reading Programme not useful?

How could the Reading Programme be improved?

What would you change? Why?

What would you keep the same? Why?

What did your friends in the class think about the Reading Programme?

How can these sessions be improved?

What do you think if I allowed you to watch English videos with English subtitles in class?

What do you think if I allowed you to browse the internet in English in class?

What other things would you prefer to read in class? Why?

What do think about the 20 minute duration?

What did your friends in the class think about the in-class silent reading sessions?

How could filling in the Reading Diary sheets be improved? Online? Why?

What did your friends in the class think about the Reading Diary sheets?

Did they change? How?

Appendix-17: In-Depth Interview Schedule [see Section-3.9.2.1] (cont.)

CONTENT QUESTION #2

WHAT DO YOU THINK ABOUT READING FOR PLEASURE?

CORE PROBES

What do you think about reading in your own Arabic?

What do you think about reading in English?

What things do you prefer doing in your free time instead of reading?

Who are the people you know who read a lot for pleasure?

POSSIBLE FOLLOW-UP PROBES

Why is reading in Arabic enjoyable / boring?

What do you read in Arabic? Why?

Why is reading in Arabic easy / difficult?

What do your friends in the class think about reading for pleasure in Arabic?

Why is reading in English enjoyable / boring?

Remember, I let you choose your books.

Why is reading in English easy / difficult? Remember I let you choose your books.

What do you now read in English? Why?

What do your friends in the class think about reading for pleasure in English?

Why?

What do they read?

What do you think of them?

What do your friends think of them?

CONTENT QUESTION #3

HOW USEFUL WAS THE READING FOR IMPROVING VOCABULARY?

CORE PROBES

What do you suggest to increase the vocabulary you learn from reading?

Should these activities be marked and added to your grade?

What do you hear your friends saying about this?

POSSIBLE FOLLOW-UP PROBES

Why?

Why?

CONTENT QUESTION #4

HOW DID YOU FEEL KNOWING THE READING WAS NOT CONNECTED TO YOUR GRADES?

CORE PROBES

Should the reading have marked study activities?

How important is the final grade once a job is found?

POSSIBLE FOLLOW-UP PROBES

Why? What kind?

Why are only marked activities important / not important?

Why is the final grade so important / not important?

Why?

What do you need to keep a job?

Why is it important / not important to get a job as soon as possible?

Appendix-17: In-Depth Interview Schedule [see Section-3.9.2.1] (cont.)

CONTENT QUESTION #5

WHAT DID YOU THINK ABOUT HAVING A FREE CHOICE OF READING?

CORE PROBES

How easy was it to find books you liked?

What were the best type of books you read?

POSSIBLE FOLLOW-UP PROBES

Why?

How much should I the teacher help to choose the books for you? Why?

Why?

Why did / didn't you read things about engineering?

How did you feel about reading things that weren't connected to engineering?

CUES FOR MORE INFORMATION

SILENT PROBES

ECHO PROMPTS

CARRY-ON FEEDBACK

REINFORCEMENT FEEDBACK

NEGATIVE REINFORCEMENT

e.g. "so you feel that..."; "so you're saying that..."

e.g. nodding gestures; "Yes", "I see"; "A-ha",
"Hmm"

e.g. "That's excellent - I never thought of that";
"That's very interesting"

e.g. "I see. Can we go back to something so I can
fully understand?"

FINAL CLOSING

Is there anything else you would like to add?

Thank-you very much for your time and comments. It is greatly appreciated.

I wish you the best in the rest of your studies.

Appendix-18: Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]

TEST-1 Based on NATION (2009)

SPELLING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	10.0	10.4	3.6	10.0	10.1	3.9	11.0	11.8	3.4	11.0	11.0	4.7	1.0	1.0
	Correctness x Certainty	7.7	8.0	3.2	8.0	7.9	3.5	9.3	9.8	3.4	8.5	9.2	4.4	1.7	0.5
	Certainty only	12.3	12.0	4.1	13.0	12.1	4.1	14.3	14.3	3.7	13.0	13.5	4.7	2.0	0.0
6000-10000wd 25 items	Correctness only	5.0	5.4	3.5	4.5	5.2	3.6	5.0	6.4	4.6	5.0	5.3	3.7	0.0	0.5
	Correctness x Certainty	2.7	3.3	2.2	2.3	3.2	2.6	3.3	4.2	3.4	2.8	3.7	3.2	0.7	0.5
	Certainty only	6.0	7.2	4.4	7.0	7.8	4.9	8.0	8.1	4.8	7.8	7.9	5.1	2.0	0.8

TEST-2 Based on NATION (2009)

TRANSLATION

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	7.0	6.8	2.8	5.5	6.6	3.6	9.0	8.7	3.1	6.5	7.4	3.7	2.0	1.0
	Correctness x Certainty	6.0	6.3	2.7	5.2	6.1	3.4	8.3	8.0	3.0	6.0	7.0	3.8	2.3	0.8
	Certainty only	8.7	9.0	3.2	8.2	8.4	3.9	10.7	11.0	3.4	9.0	9.3	4.4	2.0	0.8
6000-10000wd 25 items	Correctness only	1.0	0.8	0.9	1.0	1.0	1.9	1.0	1.4	1.4	1.0	1.6	2.1	0.0	0.0
	Correctness x Certainty	0.3	0.6	0.7	0.5	0.8	1.7	1.0	1.1	1.2	0.8	1.3	2.0	0.7	0.3
	Certainty only	2.3	2.7	2.1	2.0	2.7	3.1	2.7	3.4	2.5	2.0	3.1	3.4	0.3	0.0

TEST-3 Based on NATION (2009)

MEANING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
1000-5000wd 25 items	Correctness only	9.0	9.4	3.7	8.0	8.7	4.4	10.0	10.6	3.5	9.0	9.0	4.4	1.0	1.0
	Correctness x Certainty	7.3	7.7	3.3	6.3	7.0	4.1	8.3	8.8	3.3	6.7	7.7	4.3	1.0	0.3
	Certainty only	13.0	12.5	3.6	11.0	12.4	4.7	14.3	13.9	4.0	14.3	13.6	4.6	1.3	3.3
6000-10000wd 25 items	Correctness only	4.0	3.8	2.1	3.0	4.2	3.4	4.0	4.5	3.0	3.0	4.1	2.7	0.0	0.0
	Correctness x Certainty	2.0	2.1	1.3	2.0	2.6	2.5	2.3	2.7	2.0	2.0	2.7	2.2	0.3	0.0
	Certainty only	6.0	6.4	3.3	7.0	7.3	5.1	6.3	7.1	4.2	7.0	7.9	4.7	0.3	0.0

TEST-4 Based on SCHMITT ET AL (2001)

MEANING

		PRETEST SCORES						POSTTEST SCORES						CHANGE*	
		EXPERIMENTAL			CONTROL			EXPERIMENTAL			CONTROL			EXP.	CON.
		median	μ	σ	median	μ	σ	median	μ	σ	median	μ	σ	median	median
2000wd 30 items	Correctness only	7.0	9.0	5.7	7.0	8.6	6.5	11.0	11.3	6.2	8.5	10.1	6.9	4.0	1.5
	Correctness x Certainty	5.0	6.6	5.1	3.7	6.3	6.0	7.0	8.9	6.0	6.2	8.1	6.9	2.0	2.5
	Certainty only	13.3	13.0	5.4	11.8	13.4	6.7	15.7	15.6	5.7	16.5	16.0	6.7	2.3	4.7
3000wd 30 items	Correctness only	5.0	6.5	4.6	4.5	6.2	5.6	7.0	7.8	5.4	6.0	7.8	6.0	2.0	1.5
	Correctness x Certainty	3.7	4.6	4.1	3.0	4.6	5.0	4.3	5.9	4.9	3.5	5.7	5.9	0.7	0.5
	Certainty only	8.7	9.7	5.5	10.0	11.3	7.4	10.7	11.4	5.8	13.5	12.4	6.7	2.0	3.5
Academic wd 30 items	Correctness only	6.0	7.1	5.3	5.0	6.7	6.0	7.0	8.2	5.7	7.0	8.4	6.7	1.0	2.0
	Correctness x Certainty	3.0	5.1	4.7	3.3	5.1	5.5	4.3	6.4	5.6	4.3	6.4	6.3	1.3	1.0
	Certainty only	8.7	9.9	5.7	11.2	11.4	6.6	12.0	12.6	6.7	13.3	12.5	7.2	3.3	2.2
5000wd 30 items	Correctness only	3.0	3.4	2.9	3.0	3.6	4.5	3.0	4.0	3.2	4.0	4.8	5.5	0.0	1.0
	Correctness x Certainty	1.3	2.0	1.9	1.5	2.3	3.8	2.0	2.4	2.4	2.0	3.1	4.7	0.7	0.5
	Certainty only	6.0	6.6	4.6	6.7	7.9	5.7	8.0	8.2	5.3	9.5	9.2	6.1	2.0	2.8
10000wd 30 items	Correctness only	1.0	1.8	2.2	1.0	2.2	2.6	1.0	2.0	2.4	2.0	3.2	3.9	0.0	1.0
	Correctness x Certainty	0.7	1.0	1.3	0.7	1.1	1.8	0.3	1.0	1.5	0.8	1.7	2.6	-0.3	0.2
	Certainty only	4.0	4.7	4.4	4.8	5.4	5.1	4.0	5.3	4.8	6.0	6.9	6.0	0.0	1.2

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." Language Testing 18 (1), pp.55-88.

*Significant Change is in GREEN fill; Non-Significant Change is in RED text.

Appendix-18: Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]

(cont.)

TEST-1 Based on NATION (2009) SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.670	10.0	0.647	11.0	0.000	-3.49	0.35	1.0	10%
	Correctness x Certainty	0.751	7.7	0.779	9.3	0.000	-4.61	0.47	1.7	22%
	Certainty only	0.882	12.3	0.859	14.3	0.000	-4.54	0.46	2.0	16%
6000-10000wd 25 items	Correctness only	0.731	5.0	0.841	5.0	0.050	-1.96	0.20	0.0	0%
	Correctness x Certainty	0.684	2.7	0.853	3.3	0.023	-2.27	0.23	0.7	25%
	Certainty only	0.904	6.0	0.919	8.0	0.156	-1.42	0.14	2.0	33%

TEST-2 Based on NATION (2009) TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.717	7.0	0.736	9.0	0.000	-5.00	0.50	2.0	29%
	Correctness x Certainty	0.750	6.0	0.763	8.3	0.000	-5.08	0.51	2.3	39%
	Certainty only	0.773	8.7	0.792	10.7	0.000	-4.57	0.46	2.0	23%
6000-10000wd 25 items	Correctness only	0.194	1.0	0.423	1.0	0.003	-2.95	0.30	0.0	0%
	Correctness x Certainty	0.182	0.3	0.434	1.0	0.001	-3.41	0.34	0.7	200%
	Certainty only	0.674	2.3	0.721	2.7	0.028	-2.20	0.22	0.3	14%

TEST-3 Based on NATION (2009) MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.722	9.0	0.687	10.0	0.003	-3.01	0.30	1.0	11%
	Correctness x Certainty	0.772	7.3	0.774	8.3	0.000	-3.73	0.38	1.0	14%
	Certainty only	0.843	13.0	0.893	14.3	0.001	-3.41	0.34	1.3	10%
6000-10000wd 25 items	Correctness only	0.290	4.0	0.633	4.0	0.091	-1.69	0.17	0.0	0%
	Correctness x Certainty	0.301	2.0	0.662	2.3	0.099	-1.65	0.17	0.3	17%
	Certainty only	0.838	6.0	0.909	6.3	0.473	-0.72	0.07	0.3	6%

TEST-4 Based on SCHMITT ET AL (2001) MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.867	7.0	0.877	11.0	0.000	-4.39	0.44	4.0	57%
	Correctness x Certainty	0.907	5.0	0.918	7.0	0.000	-4.93	0.50	2.0	40%
	Certainty only	0.908	13.3	0.923	15.7	0.000	-4.25	0.43	2.3	18%
3000wd 30 items	Correctness only	0.817	5.0	0.860	7.0	0.005	-2.80	0.28	2.0	40%
	Correctness x Certainty	0.871	3.7	0.901	4.3	0.000	-3.88	0.39	0.7	18%
	Certainty only	0.920	8.7	0.928	10.7	0.012	-2.52	0.25	2.0	23%
Academic wd 30 items	Correctness only	0.867	6.0	0.878	7.0	0.029	-2.18	0.22	1.0	17%
	Correctness x Certainty	0.906	3.0	0.920	4.3	0.000	-3.54	0.36	1.3	44%
	Certainty only	0.927	8.7	0.949	12.0	0.000	-3.62	0.37	3.3	38%
5000wd 30 items	Correctness only	0.671	3.0	0.711	3.0	0.254	-1.14	0.12	0.0	0%
	Correctness x Certainty	0.716	1.3	0.780	2.0	0.063	-1.86	0.19	0.7	50%
	Certainty only	0.909	6.0	0.934	8.0	0.047	-1.99	0.20	2.0	33%
10000wd 30 items	Correctness only	0.667	1.0	0.703	1.0	0.697	-0.39	0.04	0.0	0%
	Correctness x Certainty	0.674	0.7	0.759	0.3	0.774	-0.29	0.03	-0.3	-50%
	Certainty only	0.932	4.0	0.950	4.0	0.662	-0.44	0.04	0.0	0%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill; Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill; Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{(N_{\text{pretest}} + N_{\text{posttest}})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = $((\text{posttest score} - \text{pretest score}) / \text{pretest score}) \times 100$. Note if pretest score = 0.0, then Percentage Change cannot be calculated.

Appendix-18: Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]

(cont.)

TEST-1 SPELLING

Based on NATION (2009)

EXPERIMENTAL n= 49

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	10.4	3.6	8.5	10.0	12.0	11.8	3.4	9.0	11.0	14.0
	Correctness x Certainty	8.0	3.2	5.5	7.7	10.5	9.8	3.4	7.3	9.3	12.0
	Certainty only	12.0	4.1	8.7	12.3	15.8	14.3	3.7	11.2	14.3	17.0
6000-10000wd 25 items	Correctness only	5.4	3.5	3.0	5.0	8.0	6.4	4.6	3.0	5.0	10.0
	Correctness x Certainty	3.3	2.2	1.8	2.7	4.7	4.2	3.4	1.3	3.3	6.7
	Certainty only	7.2	4.4	4.2	6.0	11.2	8.1	4.8	4.5	8.0	11.7

TEST-2 TRANSLATION

Based on NATION (2009)

EXPERIMENTAL n= 49

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	6.8	2.8	4.0	7.0	9.0	8.7	3.1	7.0	9.0	11.0
	Correctness x Certainty	6.3	2.7	4.0	6.0	8.2	8.0	3.0	5.5	8.3	10.5
	Certainty only	9.0	3.2	6.0	8.7	11.7	11.0	3.4	8.3	10.7	13.3
6000-10000wd 25 items	Correctness only	0.8	0.9	0.0	1.0	1.0	1.4	1.4	0.0	1.0	2.0
	Correctness x Certainty	0.6	0.7	0.0	0.3	1.0	1.1	1.2	0.0	1.0	1.7
	Certainty only	2.7	2.1	0.8	2.3	4.2	3.4	2.5	1.7	2.7	4.3

TEST-3 MEANING

Based on NATION (2009)

EXPERIMENTAL n= 49

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	9.4	3.7	7.0	9.0	12.0	10.6	3.5	8.0	10.0	13.0
	Correctness x Certainty	7.7	3.3	5.5	7.3	9.8	8.8	3.3	6.0	8.3	11.0
	Certainty only	12.5	3.6	9.3	13.0	15.7	13.9	4.0	11.3	14.3	16.7
6000-10000wd 25 items	Correctness only	3.8	2.1	2.0	4.0	5.0	4.5	3.0	2.0	4.0	7.0
	Correctness x Certainty	2.1	1.3	1.0	2.0	3.0	2.7	2.0	1.0	2.3	3.7
	Certainty only	6.4	3.3	4.2	6.0	9.0	7.1	4.2	4.2	6.3	9.8

TEST-4 MEANING

Based on SCHMITT ET AL (2001)

EXPERIMENTAL n= 49

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
2000wd level 30 items	Correctness only	9.0	5.7	5.0	7.0	12.5	11.3	6.2	6.5	11.0	14.5
	Correctness x Certainty	6.6	5.1	3.2	5.0	9.0	8.9	6.0	4.5	7.0	11.8
	Certainty only	13.0	5.4	9.5	13.3	16.2	15.6	5.7	11.3	15.7	20.2
3000wd level 30 items	Correctness only	6.5	4.6	3.0	5.0	9.0	7.8	5.4	4.0	7.0	11.0
	Correctness x Certainty	4.6	4.1	1.7	3.7	6.5	5.9	4.9	2.3	4.3	8.5
	Certainty only	9.7	5.5	5.3	8.7	13.2	11.4	5.8	8.3	10.7	15.2
Academic wd level 30 items	Correctness only	7.1	5.3	3.0	6.0	9.0	8.2	5.7	3.5	7.0	11.0
	Correctness x Certainty	5.1	4.7	2.0	3.0	7.2	6.4	5.6	2.2	4.3	8.3
	Certainty only	9.9	5.7	5.3	8.7	14.8	12.6	6.7	7.8	12.0	17.5
5000wd level 30 items	Correctness only	3.4	2.9	1.0	3.0	5.0	4.0	3.2	2.0	3.0	5.5
	Correctness x Certainty	2.0	1.9	0.3	1.3	3.0	2.4	2.4	0.8	2.0	3.0
	Certainty only	6.6	4.6	2.3	6.0	9.3	8.2	5.3	4.0	8.0	11.5
10000wd level 30 items	Correctness only	1.8	2.2	0.0	1.0	3.0	2.0	2.4	0.0	1.0	3.0
	Correctness x Certainty	1.0	1.3	0.0	0.7	1.2	1.0	1.5	0.0	0.3	1.7
	Certainty only	4.7	4.4	1.0	4.0	7.8	5.3	4.8	1.0	4.0	8.2

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

Appendix-18: Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]

(cont.)

TEST-1 Based on NATION (2009)
SPELLING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.718	10.0	0.826	11.0	0.181	-1.34	0.17	1.0	10%
	Correctness x Certainty	0.784	8.0	0.868	8.5	0.060	-1.88	0.24	0.5	6%
	Certainty only	0.878	13.0	0.909	13.0	0.049	-1.97	0.25	0.0	0%
6000-10000wd 25 items	Correctness only	0.747	4.5	0.765	5.0	0.898	-0.13	0.02	0.5	11%
	Correctness x Certainty	0.790	2.3	0.832	2.8	0.087	-1.71	0.22	0.5	21%
	Certainty only	0.923	7.0	0.923	7.8	0.984	-0.02	0.00	0.8	12%

TEST-2 Based on NATION (2009)
TRANSLATION

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.829	5.5	0.829	6.5	0.003	-2.96	0.38	1.0	18%
	Correctness x Certainty	0.835	5.2	0.854	6.0	0.002	-3.05	0.39	0.8	16%
	Certainty only	0.851	8.2	0.877	9.0	0.043	-2.03	0.26	0.8	10%
6000-10000wd 25 items	Correctness only	0.766	1.0	0.744	1.0	0.008	-2.64	0.34	0.0	0%
	Correctness x Certainty	0.804	0.5	0.771	0.8	0.004	-2.87	0.37	0.3	67%
	Certainty only	0.853	2.0	0.860	2.0	0.116	-1.57	0.20	0.0	0%

TEST-3 Based on NATION (2009)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
1000-5000wd 25 items	Correctness only	0.809	8.0	0.809	9.0	0.635	-0.48	0.06	1.0	13%
	Correctness x Certainty	0.867	6.3	0.861	6.7	0.121	-1.55	0.20	0.3	5%
	Certainty only	0.916	11.0	0.910	14.3	0.036	-2.09	0.27	3.3	30%
6000-10000wd 25 items	Correctness only	0.731	3.0	0.578	3.0	0.726	-0.35	0.05	0.0	0%
	Correctness x Certainty	0.779	2.0	0.697	2.0	0.943	-0.07	0.01	0.0	0%
	Certainty only	0.940	7.0	0.917	7.0	0.339	-0.96	0.13	0.0	0%

TEST-4 Based on SCHMITT ET AL (2001)
MEANING

		Pretest Internal Consistency		Posttest Internal Consistency		p (<0.05) for significant change**	Z** value	r Effect Size***	Change** in Median (no. of words)	Change** in Median (%****)
		Cronbach's alpha (>0.7*)	Median	Cronbach's alpha (>0.7*)	Median					
2000wd 30 items	Correctness only	0.901	7.0	0.907	8.5	0.006	-2.76	0.35	1.5	21%
	Correctness x Certainty	0.938	3.7	0.941	6.2	0.000	-3.51	0.44	2.5	68%
	Certainty only	0.952	11.8	0.949	16.5	0.002	-3.05	0.38	4.7	39%
3000wd 30 items	Correctness only	0.881	4.5	0.885	6.0	0.000	-3.51	0.44	1.5	33%
	Correctness x Certainty	0.915	3.0	0.934	3.5	0.001	-3.29	0.41	0.5	17%
	Certainty only	0.961	10.0	0.953	13.5	0.102	-1.64	0.20	3.5	35%
Academic wd 30 items	Correctness only	0.897	5.0	0.907	7.0	0.002	-3.08	0.38	2.0	40%
	Correctness x Certainty	0.930	3.3	0.939	4.3	0.004	-2.85	0.36	1.0	30%
	Certainty only	0.946	11.2	0.960	13.3	0.189	-1.31	0.16	2.2	19%
5000wd 30 items	Correctness only	0.881	3.0	0.906	4.0	0.017	-2.40	0.30	1.0	33%
	Correctness x Certainty	0.930	1.5	0.945	2.0	0.020	-2.33	0.29	0.5	33%
	Certainty only	0.941	6.7	0.951	9.5	0.150	-1.44	0.18	2.8	42%
10000wd 30 items	Correctness only	0.736	1.0	0.836	2.0	0.117	-1.57	0.20	1.0	100%
	Correctness x Certainty	0.822	0.7	0.881	0.8	0.067	-1.83	0.23	0.2	25%
	Certainty only	0.956	4.8	0.967	6.0	0.098	-1.66	0.21	1.2	24%

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:

<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

*See Pallant (2010:97):

Good Internal Consistency shown by GREEN fill; Poor Internal Consistency shown by RED text.

**Wilcoxon signed-rank test for non-normally distributed repeated measures. See Pallant (2010:230): Significant Change - GREEN fill; Non-Significant - RED text.

***Effect Size $r = |Z| / \sqrt{N(\text{pretest} + N\text{posttest})}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:232).

****Percentage (%) Change = ((posttest score - pretest score) / pretest score) x 100. Note if pretest score = 0, then Percentage Change cannot be calculated.

Note: Less participants than the total control cohort of 36 participants are recorded, because a few of them were absent for some of the posttests, with these absentees being different individuals in different tests.

Appendix-18: Results of the Stage-3 Main Study Vocabulary Tests [see Section-4.2]

(cont.)

TEST-1 SPELLING

Based on NATION (2009)

CONTROL n= 30

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	10.1	3.9	6.8	10.0	12.0	11.0	4.7	9.0	11.0	13.0
	Correctness x Certainty	7.9	3.5	5.5	8.0	9.1	9.2	4.4	6.8	8.5	11.1
	Certainty only	12.1	4.1	8.8	13.0	14.7	13.5	4.7	10.8	13.0	16.0
6000-10000wd 25 items	Correctness only	5.2	3.6	2.0	4.5	7.5	5.3	3.7	3.0	5.0	7.0
	Correctness x Certainty	3.2	2.6	1.2	2.3	4.7	3.7	3.2	1.7	2.8	5.1
	Certainty only	7.8	4.9	3.0	7.0	11.2	7.9	5.1	3.8	7.8	10.8

TEST-2 TRANSLATION

Based on NATION (2009)

CONTROL n= 30

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	6.6	3.6	4.8	5.5	9.0	7.4	3.7	5.0	6.5	9.0
	Correctness x Certainty	6.1	3.4	4.0	5.2	8.1	7.0	3.8	4.2	6.0	9.0
	Certainty only	8.4	3.9	5.3	8.2	9.8	9.3	4.4	5.7	9.0	11.8
6000-10000wd 25 items	Correctness only	1.0	1.9	0.0	1.0	1.0	1.6	2.1	0.0	1.0	2.0
	Correctness x Certainty	0.8	1.7	0.0	0.5	1.0	1.3	2.0	0.0	0.8	1.8
	Certainty only	2.7	3.1	0.7	2.0	3.4	3.1	3.4	1.3	2.0	3.4

TEST-3 MEANING

Based on NATION (2009)

CONTROL n= 29

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
1000-5000wd 25 items	Correctness only	8.7	4.4	6.0	8.0	11.0	9.0	4.4	6.0	9.0	11.0
	Correctness x Certainty	7.0	4.1	4.0	6.3	8.3	7.7	4.3	5.2	6.7	10.0
	Certainty only	12.4	4.7	9.5	11.0	16.2	13.6	4.6	10.0	14.3	16.7
6000-10000wd 25 items	Correctness only	4.2	3.4	1.5	3.0	6.0	4.1	2.7	2.5	3.0	5.5
	Correctness x Certainty	2.6	2.5	0.7	2.0	4.0	2.7	2.2	1.3	2.0	3.2
	Certainty only	7.3	5.1	3.0	7.0	11.0	7.9	4.7	3.5	7.0	11.8

TEST-4 MEANING

Based on SCHMITT ET AL (2001)

CONTROL n= 32

		Pretest Descriptive Statistics					Posttest Descriptive Statistics				
		Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile	Mean	Standard Deviation	Lower Quartile	Median	Upper Quartile
2000wd level 30 items	Correctness only	8.6	6.5	5.0	7.0	10.8	10.1	6.9	5.3	8.5	13.0
	Correctness x Certainty	6.3	6.0	3.0	3.7	8.5	8.1	6.9	3.1	6.2	11.6
	Certainty only	13.4	6.7	8.8	11.8	17.2	16.0	6.7	10.8	16.5	20.7
3000wd level 30 items	Correctness only	6.2	5.6	3.0	4.5	8.0	7.8	6.0	4.0	6.0	9.8
	Correctness x Certainty	4.6	5.0	1.3	3.0	6.7	5.7	5.9	2.7	3.5	7.5
	Certainty only	11.3	7.4	6.4	10.0	17.2	12.4	6.7	5.9	13.5	16.7
Academic wd level 30 items	Correctness only	6.7	6.0	3.0	5.0	8.0	8.4	6.7	3.3	7.0	10.0
	Correctness x Certainty	5.1	5.5	2.0	3.3	6.5	6.4	6.3	2.3	4.3	7.7
	Certainty only	11.4	6.6	6.7	11.2	16.8	12.5	7.2	6.4	13.3	16.3
5000wd level 30 items	Correctness only	3.6	4.5	1.0	3.0	4.0	4.8	5.5	1.0	4.0	6.0
	Correctness x Certainty	2.3	3.8	0.8	1.5	2.3	3.1	4.7	1.0	2.0	3.3
	Certainty only	7.9	5.7	3.6	6.7	12.3	9.2	6.1	3.7	9.5	13.2
10000wd level 30 items	Correctness only	2.2	2.6	0.0	1.0	3.8	3.2	3.9	0.3	2.0	4.8
	Correctness x Certainty	1.1	1.8	0.0	0.7	1.7	1.7	2.6	0.1	0.8	2.2
	Certainty only	5.4	5.1	0.8	4.8	8.2	6.9	6.0	0.9	6.0	10.2

- Nation, I.S.P. (2009). "Vocabulary size test." Available from:
<http://www.victoria.ac.nz/lals/staff/paul-nation/vocabulary%20size%20test.pdf> [accessed 27/4/2009].

- Schmitt, N., Schmitt, D. & Clapham, C. (2001). "Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test." *Language Testing* 18 (1), pp.55-88.

Note: Less participants than the total control cohort of 36 participants are recorded, because a few of them were absent for some of the posttests, with these absentees being different individuals in different tests.

Appendix-19: Results of End-of-Semester English Course Exam [see Section-4.3]

END-OF-SEMESTER
ENGLISH COURSE EXAMS

	STAGE-2 MAIN PILOT		STAGE-3 MAIN STUDY		
	Main Pilot Cohort	All Students in Batch	Experimental Cohort	Control Cohort	All Students in Batch
Number of Participants in the Study	52	52	49	36	85
Number who took the Exam	52	730	49	36	691
Mean (%)	83	78	80	71	72
Standard Deviation (%)	13	16	12	21	18
Minimum (%)	47	20	51	28	18
Lower Quartile Q1 (%)	77	70	70	57	60
Median Q2 (%)	87	82	83	78	75
Upper Quartile Q3 (%)	92	90	90	90	86
Maximum (%)	98	100	99	99	99
Range (%)	51	80	48	71	81

COMPARING GROUPS:	p (<0.05 for significant* difference in results)	U* value	Z* value	r Effect Size**
Main Pilot Cohort & All Students in Batch	0.025	15458	-2.24	0.08
Experimental Cohort & Control Cohort	0.101	698	-1.64	0.18
Experimental Cohort & All Students in Batch	0.002	12490	-3.07	0.12
Control Cohort & All Students in Batch	0.835	12182	-0.21	0.01

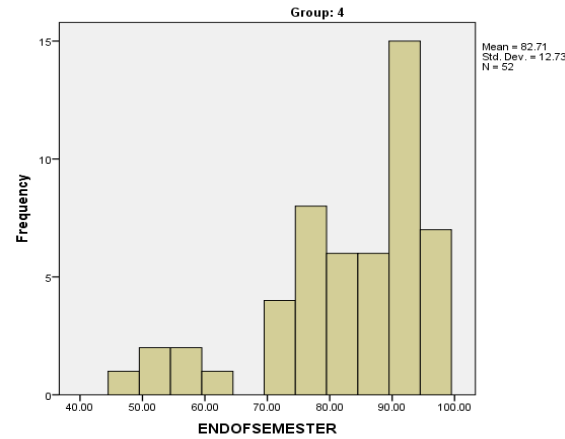
*Mann-Whitney U Test for non-normally distributed independent samples. See Pallant (2010:227-230).

**Effect Size $r = |Z| / \sqrt{N_{total}}$; $r=0.1$ small; $r=0.3$ medium; $r=0.5$ large. See Pallant (2010:230).

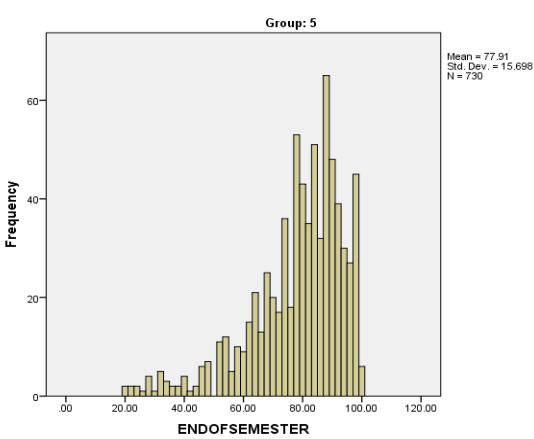
Appendix-19: Results of End-of-Semester English Course Exam [see Section-4.3]
(cont.)

STAGE-2 MAIN PILOT

Main Pilot Cohort

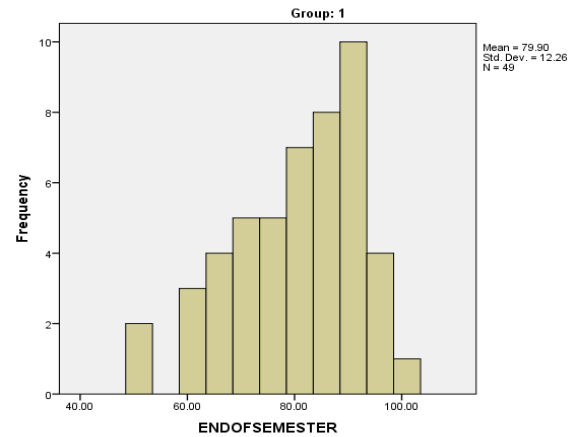


All Students in Batch

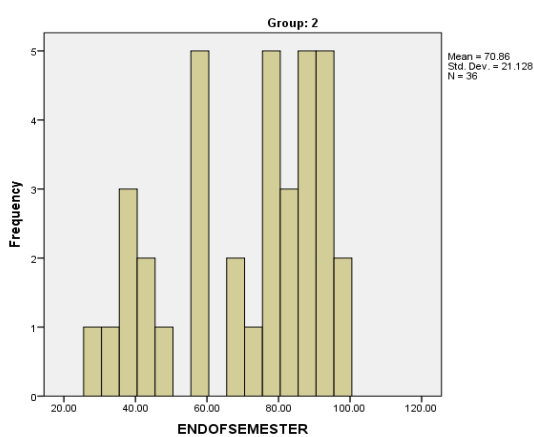


STAGE-3 MAIN STUDY

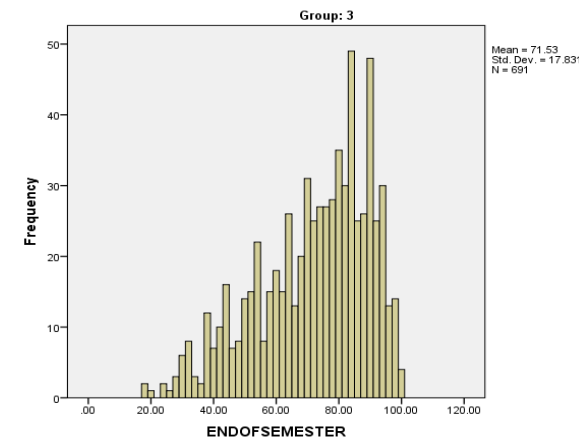
Experimental Cohort



Control Cohort



All Students in Batch



Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4]

READING DIARY SHEETS - Types of Reading Material

ITEM-3 (Type of Reading Material) - DETAILED Options	
1	College/Work Book
2	Teacher's Lecture Notes
3	Student's Own Notes
4	Book with Translation
5	Graded Reader
6	Simplified Book with Pictures
7	Simplified Book with FEW Pictures
8	Newspaper
9	Magazine
10	Comic/Movie Story-Book
11	Full Novel
12	Internet Text
13	Computer/Mobile Text Offline
14	English Video + English Subtitles
15	English Video + Arabic Subtitles
16	Arabic Video + English Subtitles
17	Arabic Video + Arabic Subtitles
18	Other - instruction leaflets, user guides, manuals, catalogues, tourist guides, tickets, menus, posters, and other materials that did not fit into the other categories such as books with many pictures but with unsimplified language, like books aimed for young people on wildlife, science and sport

ITEM-3 (Type of Reading Material) - OVERALL Options		
1	College Material	Detailed Option-1,2,3
2	Simplified Book Material	Detailed Option-4,5,6,7
3	Unsimplified Book Material	Detailed Option-8,9,10,11,18
4	Screen Text with No Video	Detailed Option-12,13
5	Video with Subtitles	Detailed Option-14,15,16,17

Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

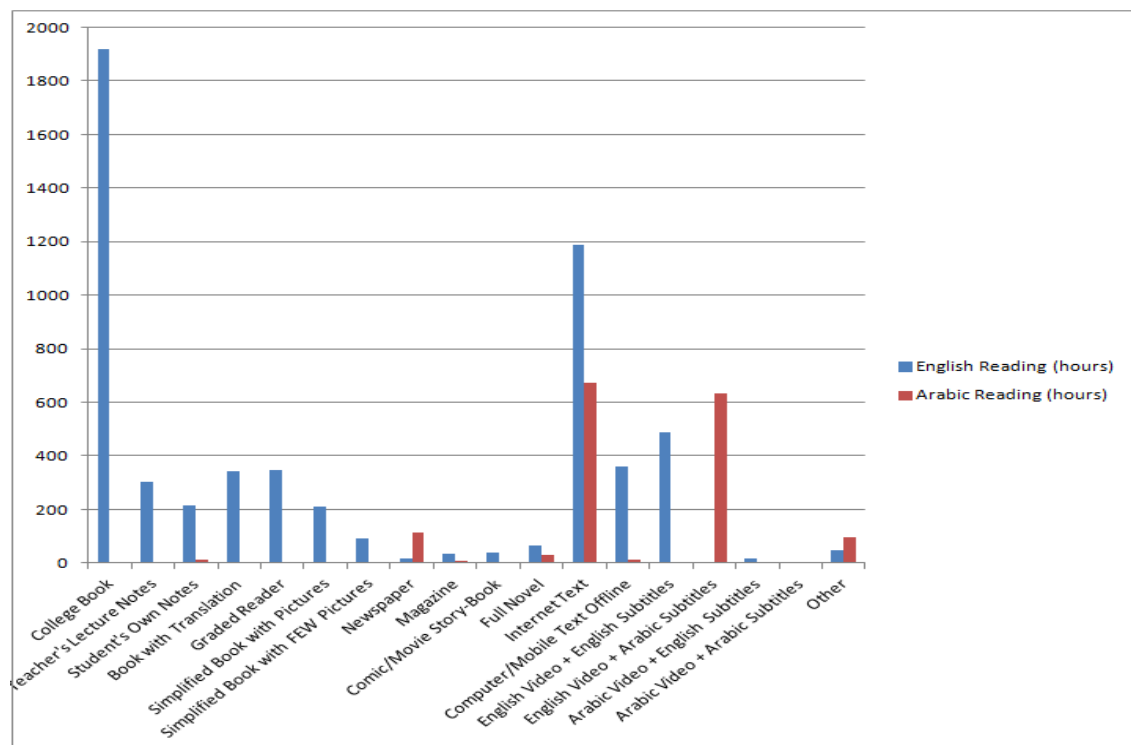
READING DIARY SHEETS - Topics of Reading

ITEM-4 (Topic of Reading Material) - Formal Education; Informal Learning; Information; Communication; Entertainment	
1	College-Related - <i>College Textbooks & Workbooks; Lecture Notes; Students' Notes; Assignments; English Language Computer Lab; etc.</i>
2	Children's Stories - <i>Fairy Tales</i> e.g. <i>Snow White, Cinderella, Little Mermaid, Beauty and the Beast, Sleeping Beauty, Jack & Beanstalk, Pinocchio, Elves & Shoemaker, Hansel and Gretel, Ugly Duckling, Puss in Boots, etc.</i> ; - <i>Educational</i> e.g. <i>Ladybird Key Words (Peter & Jane), Sesame Street</i> ; - <i>Cartoon-based</i> e.g. <i>Bugs Bunny, Scooby Doo, Mickey Mouse, Ben 10, Mr. Men, Shrek, Paddington Bear, etc.</i> ; - <i>Aesop's Fables</i> ; - <i>Short Stories</i> e.g. <i>Big Pancake, Selfish Giant, Wooden Toy, Magic Horse, Wicked Prince, King & Falcon, Big Pancake, etc.</i> ; - <i>Animal Stories</i> e.g. <i>Good Morning Gorilla, Polar Bears Past Bedtime</i> ; - <i>Longer Stories</i> e.g. <i>Wizard of Oz, Lord Fauntleroy, Black Beauty, Secret Garden, A Little Princess, Railway Children, etc.</i>
3	Action/Adventure/Thriller - <i>Adventures</i> e.g. <i>Aladdin, Tom Sawyer, Ali Baba, Treasure Island, Jungle Book, Round World in 80 days, Old Man and Sea, Prince & Pauper, Robin Hood, Sindbad, Robinson Crusoe, Oliver Twist, Gulliver's Travels, 20000 Leagues Under Sea, White Fang, Call of the Wild, Lost World, King Solomon's Mines, etc.</i> ; - <i>Action</i> e.g. <i>Terminator, King Arthur, James Bond, Spiderman, Ironman, Zorro, Kidnapped, Mission Impossible, Three Musketeers, Harry Potter, Pirates of the Caribbean, Assassin's Creed, etc.</i> ; - <i>Thrillers</i> e.g. <i>Godfather, President's Murderer, One Way Ticket, Dead Man's Island, Death in the Freezer, The Count of Monte Cristo, White Collar, The Mentalist, etc.</i> ; - <i>Science Fiction</i> e.g. <i>Frankenstein, etc.</i> ; - <i>Detective</i> e.g. <i>Sherlock Holmes, Agatha Christie, Nancy Drew, etc.</i> ; - <i>Horror</i> e.g. <i>Saw, Ghost Stories, Dracula, etc.</i>
4	<i>Texts/Emails - Texts; Emails; Social Media e.g. Facebook, Twitter, etc.</i>
5	<i>News - (but not sport); Current Affairs; Breakfast TV; Interviews</i>
6	English Language Learning - <i>Websites, Applications, CDs, Software</i> ; - <i>Videos</i> e.g. <i>Abu Omar, Mr. Duncan</i> ; - <i>English Language Textbooks [unrelated to the College]</i> ; - <i>Reading when Teaching English; Bilingual Dictionaries</i>
7	Dramas/Soap Operas - <i>Soap Operas, Dramas, Series & Romance</i> e.g. <i>Prison Break, Dexter, Great Expectations, The Event, Leverage, Friends, Outcasts, David Copperfield, Grey's Anatomy, Touch, Beverly Hills 90210, Castle, East of Eden, Loser Takes All, Lost, Jane Eyre, Oliver Twist, Stories from the Five Towns, Lord Jim, Lottery Winner, etc.</i>
8	Religion - <i>Islam; Quran</i>
9	Sports/Games - <i>Sport; Sports News; Video Games</i>
10	Documentaries/Information - <i>Documentaries & History</i> e.g. <i>Sinking of the Titanic, etc.</i> - <i>Biographies</i> e.g. <i>Elephant Man, etc.</i> - <i>Information</i> e.g. <i>Art, Perfume, Guinness Book of Records, Interior Design, Photography, www.ehow.com, etc.</i> - <i>Geography</i> e.g. <i>National Geographic, Disasters, Wildlife, etc.</i> - <i>Science & Health</i> - <i>Travel & Booking Tickets</i>
11	Plays e.g. <i>Merchant of Venice; Macbeth; Midsummer Night Dream</i>
12	Technology - <i>Instructions</i> e.g. <i>for a new device, Windows installation, GPS in the car</i> - <i>Buying Technical Items</i> e.g. <i>cars, computers, cameras, EBay, etc.</i> - <i>Engineering jobs and careers</i>
13	General Browsing - <i>General Browsing; Wikipedia</i>
14	Comedies e.g. <i>Diary of a Wimpy Kid, Centerville Ghost, Raising Hope, How I Met Your Mother, Three Men in a Boat, Two and a Half Men</i>
15	Self-Improvement - <i>Behaviour; Learning; Success</i> e.g. <i>Rich Man Poor Dad</i>
16	Political Stories e.g. <i>Animal Farm</i>

Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

Item-3 (Type of Reading Material) – DETAILED OPTIONS

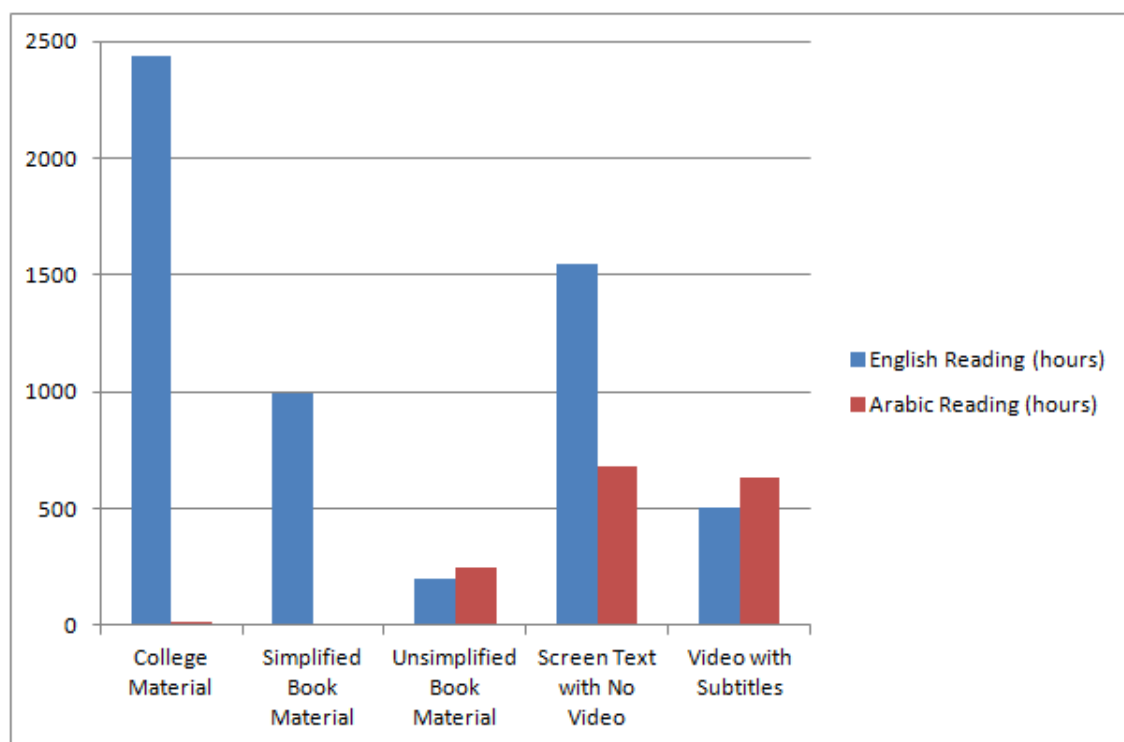
Sum of DURATION OF READING (hrs)	Column Labels		
Row Labels	English Reading (hours)	Arabic Reading (hours)	Grand Total
College Book	1917		1917
Teacher's Lecture Notes	301		301
Student's Own Notes	216	11	227
Book with Translation	343	1	344
Graded Reader	347		347
Simplified Book with Pictures	210	0	210
Simplified Book with FEW Pictures	92	1	93
Newspaper	15	114	129
Magazine	33	7	40
Comic/Movie Story-Book	39		39
Full Novel	66	31	97
Internet Text	1189	671	1860
Computer/Mobile Text Offline	359	13	372
English Video + English Subtitles	487		487
English Video + Arabic Subtitles		632	632
Arabic Video + English Subtitles	16		16
Arabic Video + Arabic Subtitles		4	4
Other	49	96	144
Grand Total	5676	1580	7256



Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

Item-3 (Type of Reading Material) – OVERALL OPTIONS

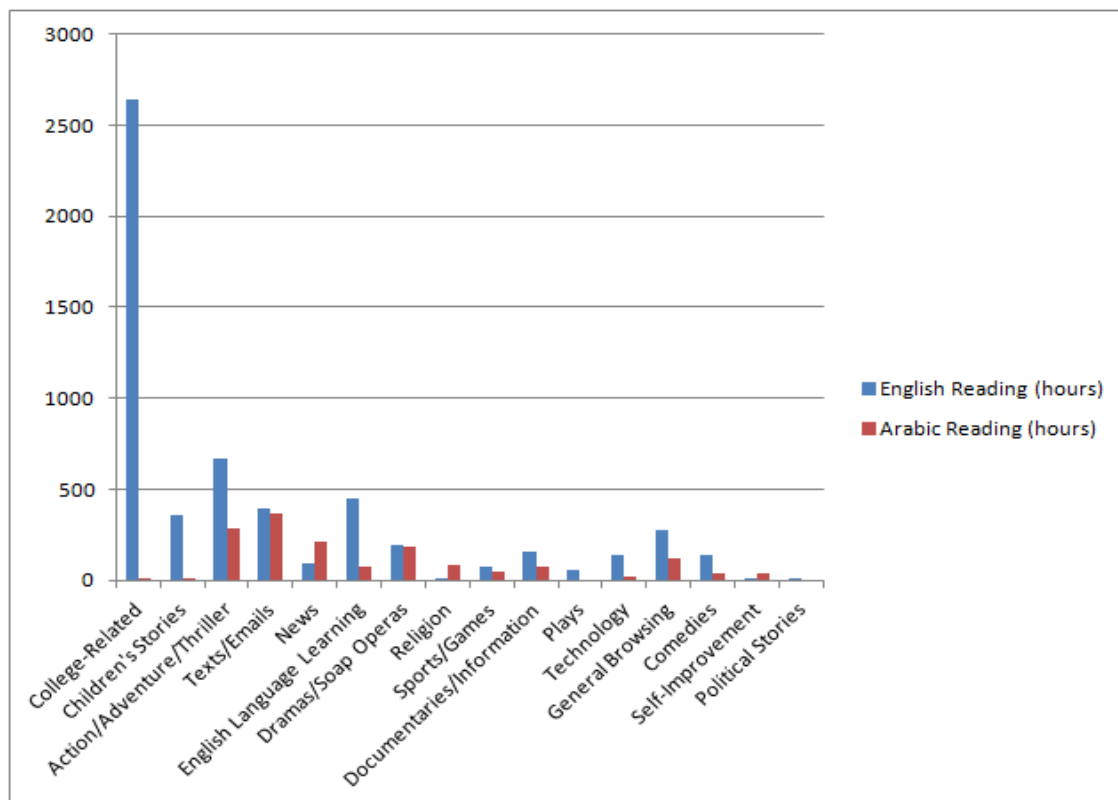
Sum of DURATION OF READING (hrs)	Column Labels		
Row Labels	English Reading (hours)	Arabic Reading (hours)	Grand Total
College Material	2433	11	2444
Simplified Book Material	991	2	994
Unsimplified Book Material	202	247	449
Screen Text with No Video	1548	684	2232
Video with Subtitles	502	635	1138
Grand Total	5676	1580	7256



Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

Item-4 (Topic of Reading Material)

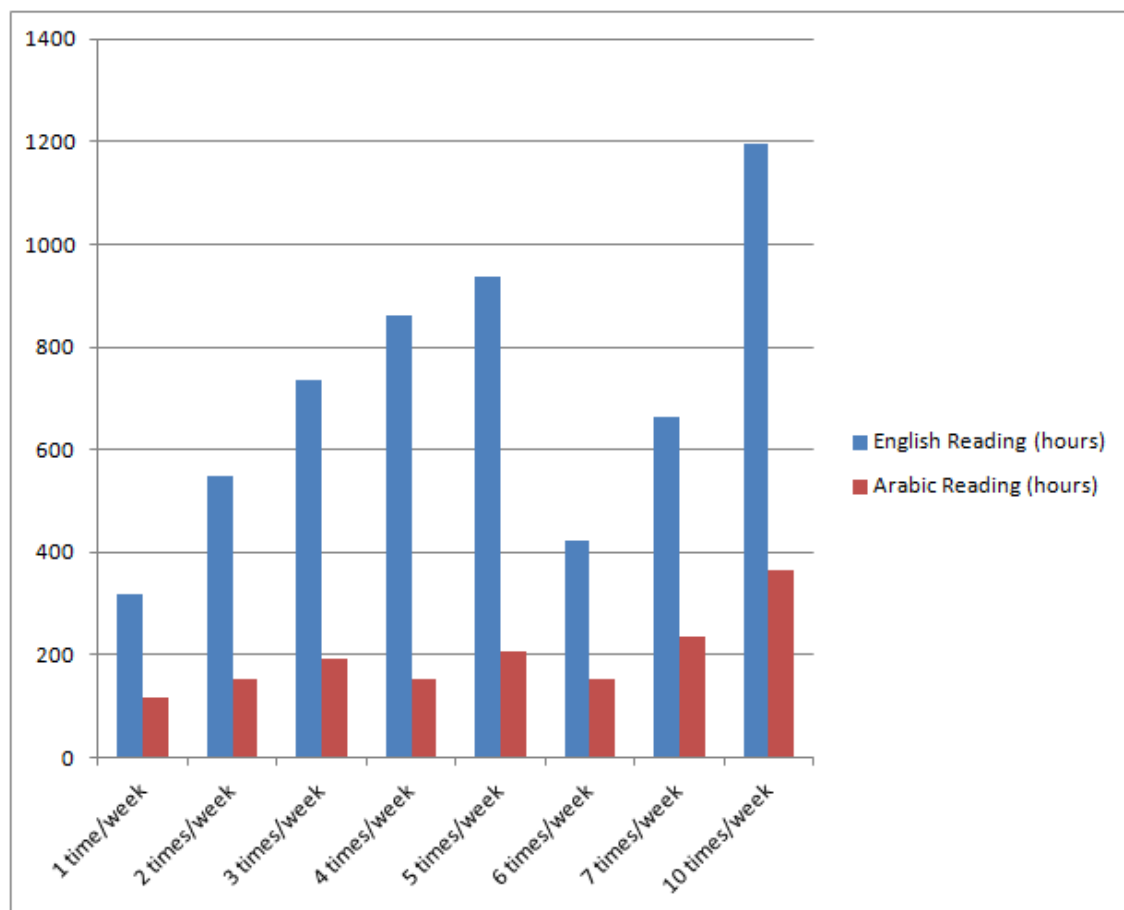
Sum of DURATION OF READING (hrs)	Column Labels		
Row Labels	English Reading (hours)	Arabic Reading (hours)	Grand Total
College-Related	2639	11	2651
Children's Stories	355	7	362
Action/Adventure/Thriller	666	289	955
Texts/Emails	394	364	758
News	94	210	303
English Language Learning	448	79	527
Dramas/Soap Operas	199	189	388
Religion	15	89	103
Sports/Games	78	48	127
Documentaries/Information	162	73	235
Plays	62		62
Technology	137	19	156
General Browsing	277	122	399
Comedies	139	40	179
Self-Improvement	3	41	44
Political Stories	8		8
Grand Total	5676	1580	7256



Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

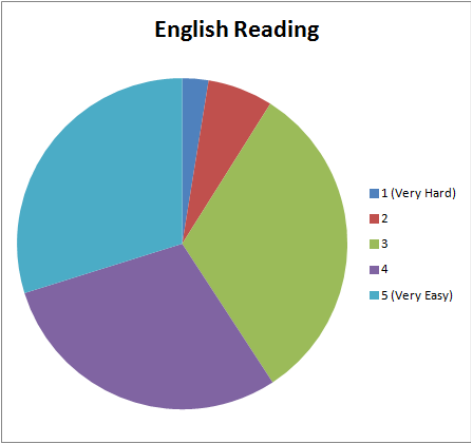
Item-6 (Frequency of Reading)

Sum of DURATION OF READING (hrs)	Column Labels	
Row Labels	English Reading (hours)	Arabic Reading (hours)
1 time/week	320	117
2 times/week	547	154
3 times/week	735	193
4 times/week	859	154
5 times/week	936	206
6 times/week	422	153
7 times/week	662	237
10 times/week	1195	367
Grand Total	5676	1580

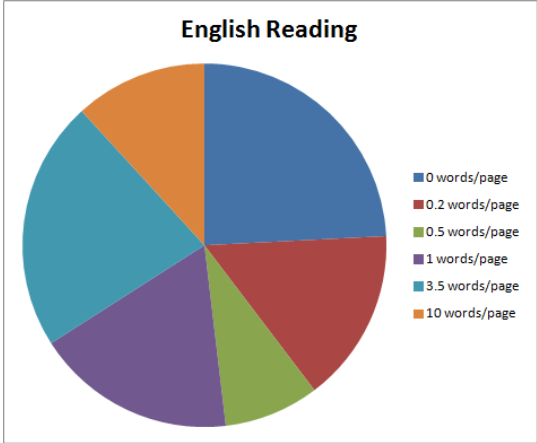


Appendix-20: Results of the Reading Diaries for the Experimental Cohort (n=49) [see Section-4.4] (cont.)

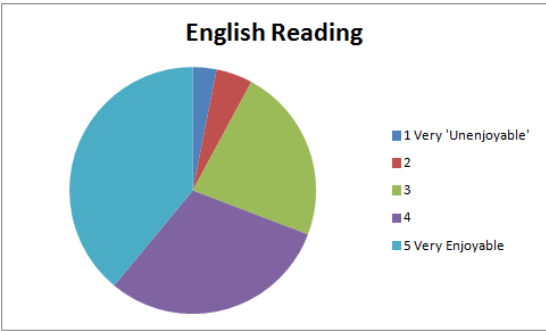
Item-7 (Ease of Reading)



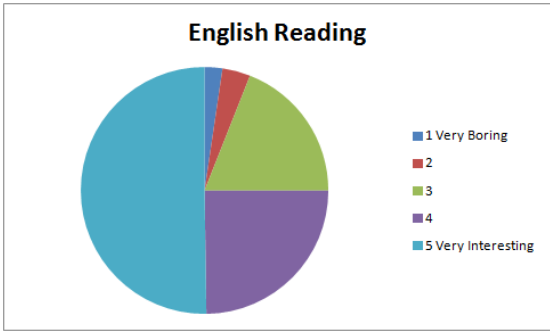
Item-8 (Dictionary Use)



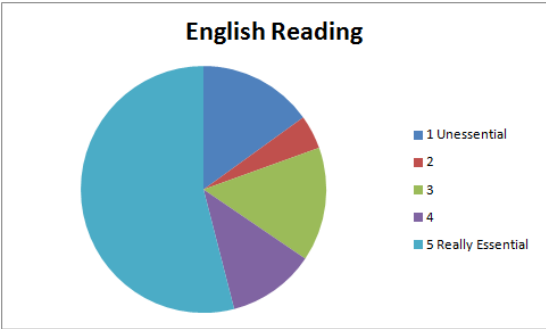
Item-9 (Enjoyment)



Item-10 (Interest)



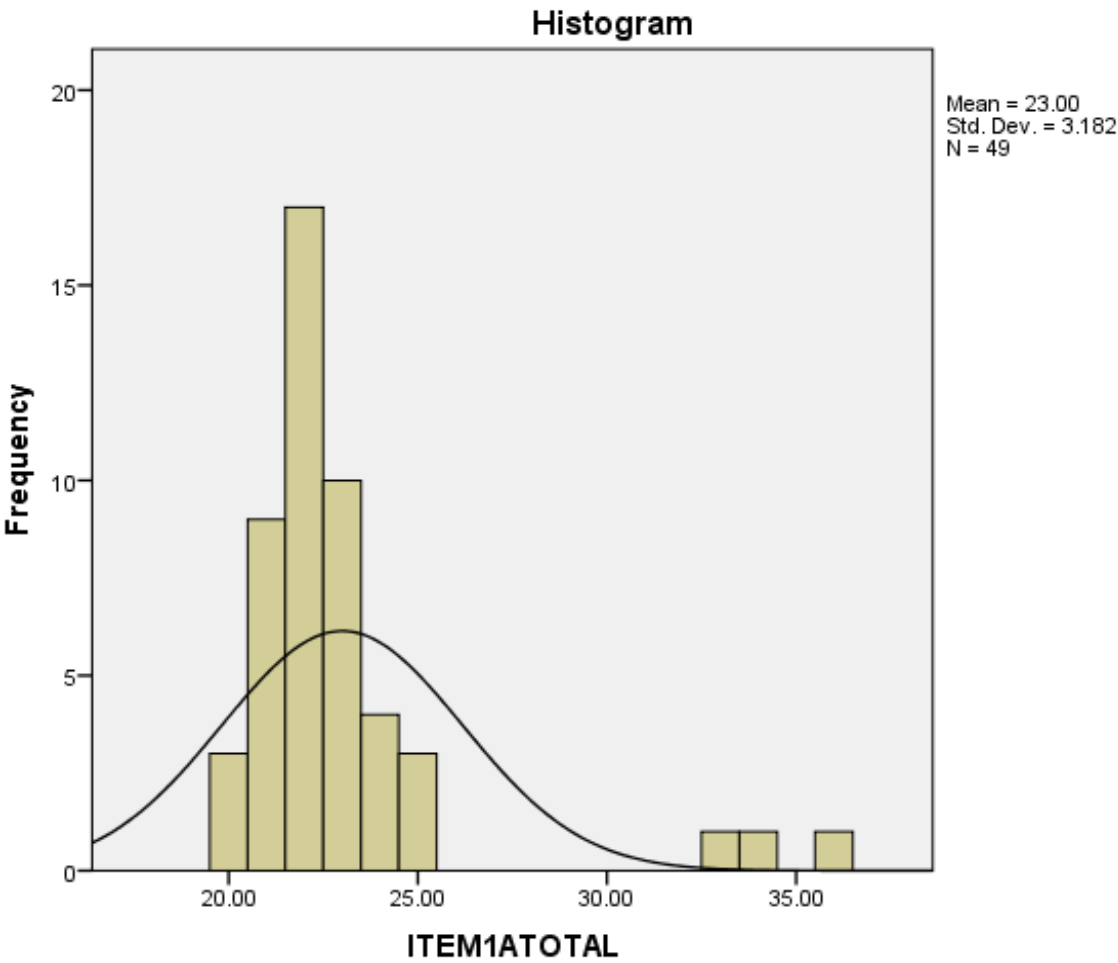
Item-11 (How Essential the Reading was)



Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1]

ITEM 1
AGE

		n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
ALL STUDENTS	years	49	23	22	22	23	20	36	3

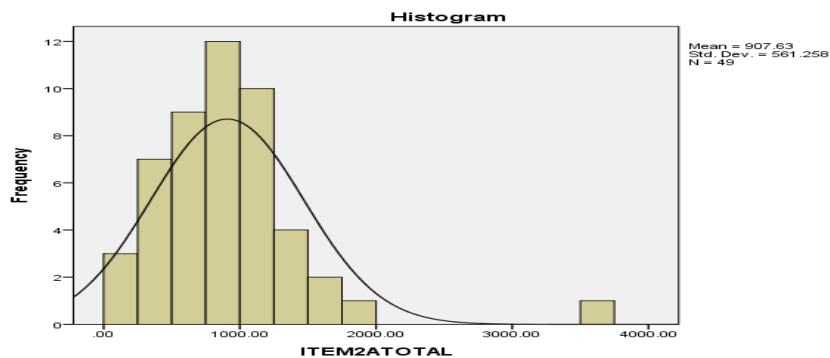
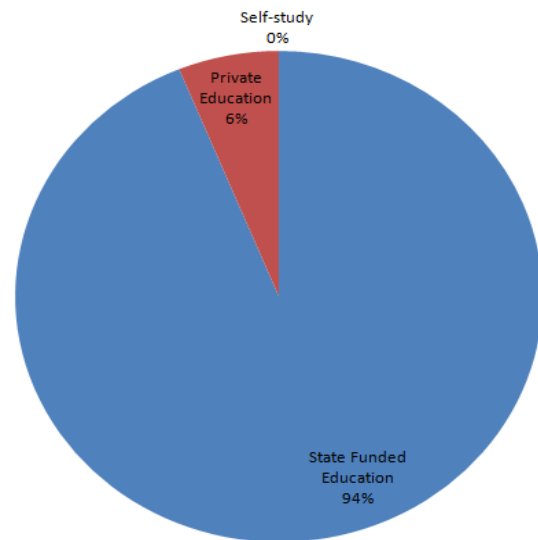
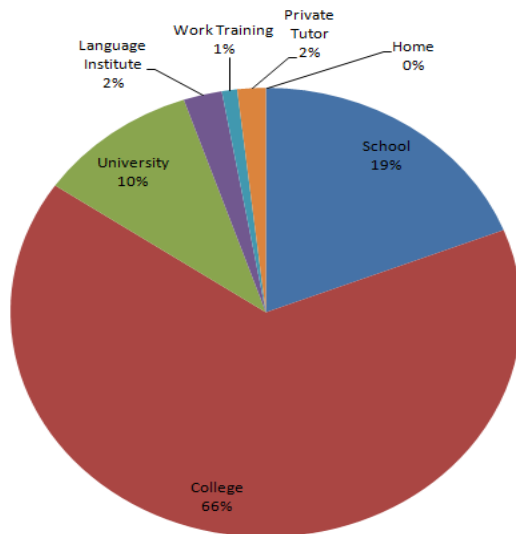


Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1] (cont.)

ITEM 2

PRIOR ENGLISH STUDIES FACTORED BY PROPORTION OF TEACHING IN ENGLISH

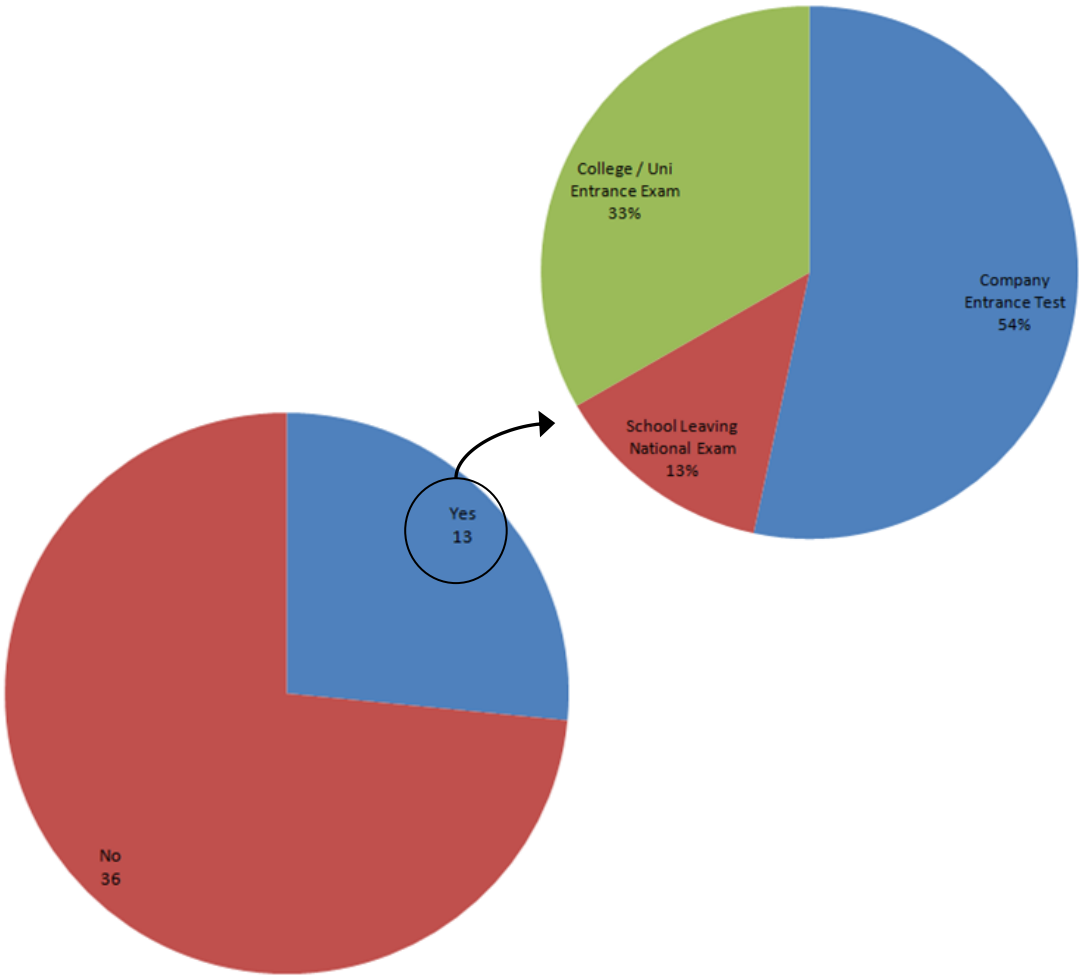
			n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
A. LEVEL	School	hours	49	173	59	126	252	0	720	155
	College	hours	49	595	480	618	780	75	1050	253
	University	hours	49	93	0	0	0	0	2700	415
	Language Institute	hours	49	22	0	0	0	0	752	108
	Work/Training	hours	49	9	0	0	0	0	390	56
	Private Tutor	hours	49	16	0	0	0	0	784	112
	Home Self-study	hours	49	0	0	0	0	0	5	1
	ALL LEVELS	hours	49	908	642	839	1131	122	3543	561
B. TYPE	State Funded Education	hours	49	851	576	828	1104	92	3543	543
	Private Education	hours	49	56	0	0	0	0	784	185
	Self-study	hours	49	0	0	0	0	0	5	1
	ALL TYPES	hours	49	908	642	839	1131	122	3543	561
C. PLACE	Saudi Arabia	hours	49	900	642	839	1131	122	3543	552
	Other (Far East)	hours	49	8	0	0	0	0	390	56
	ALL PLACES	hours	49	908	642	839	1131	122	3543	561



Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1] (cont.)

ITEM 3
EXTERNAL ENGLISH EXAMS

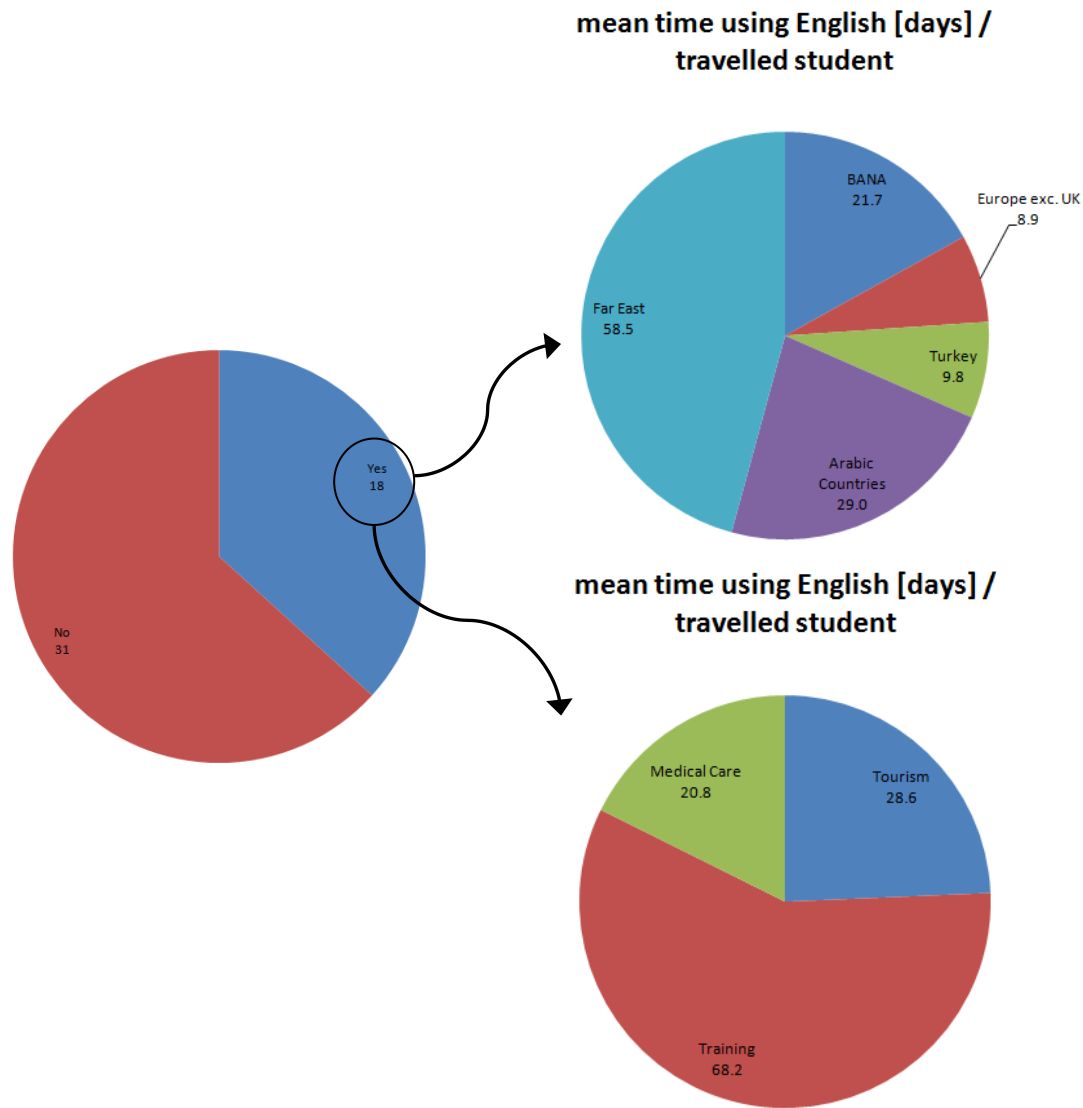
			n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
A. TYPE	Company Entrance Test	% score	8	74	67	74	80	50	100	15
	School Leaving National Exam	% score	2	59	54	59	63	50	67	12
	College / Uni Entrance Exam	% score	5	69	67	70	86	35	87	21



Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1] (cont.)

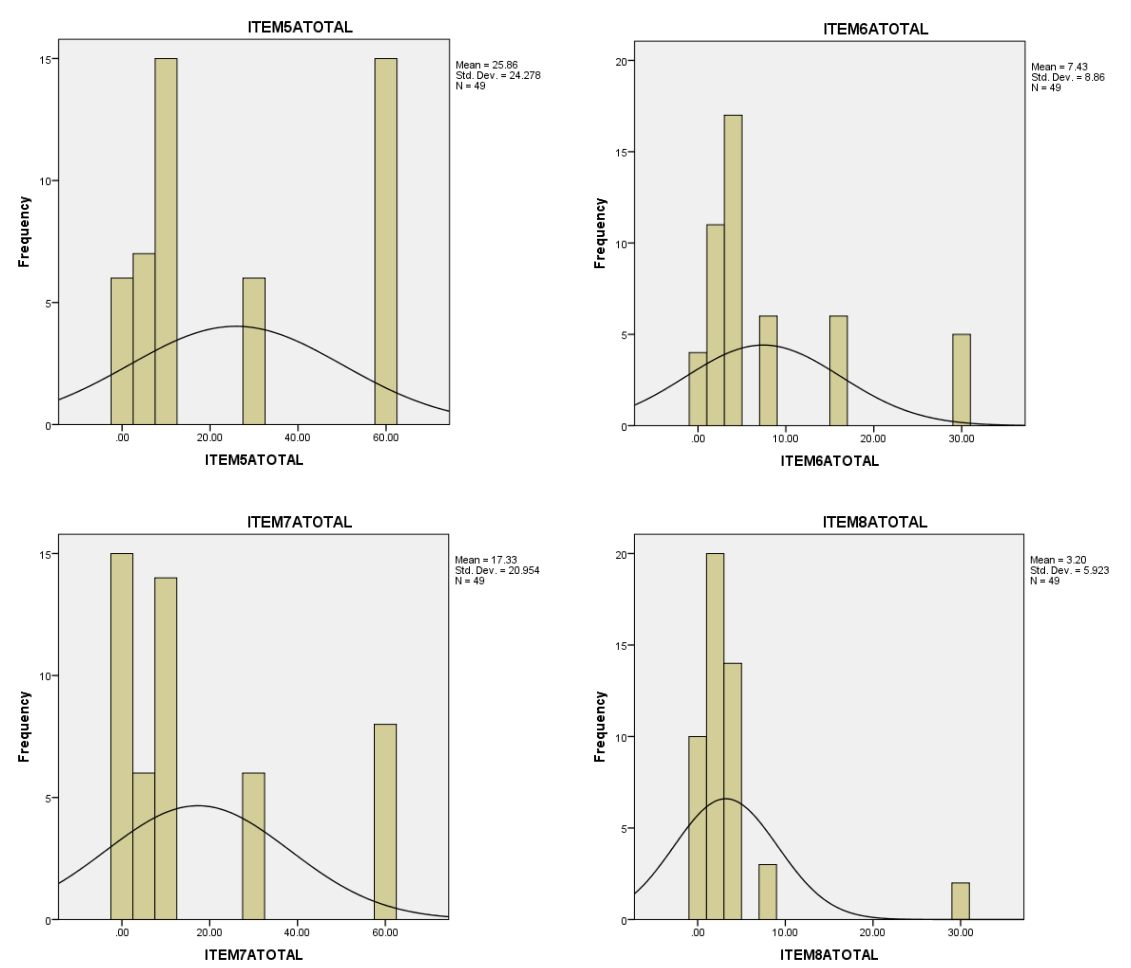
ITEM 4
TRAVEL ABROAD FACTORED BY PROPORTION OF TIME USING ENGLISH

			n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
A. PLACE	BANA	days	3	22	15	27	32	2	36	18
	Europe exc. UK	days	2	9	5	9	13	1	17	11
	Turkey	days	1	10	10	10	10	10	10	N/A
	Arabic Countries	days	14	29	3	6	18	0	288	75
	Far East	days	4	59	19	56	95	12	111	49
	ALL who travelled	days	18	41	2	6	35	0	317	79
B. PURPOSE	Tourism	days	17	29	2	6	16	0	317	75
	Training	days	3	68	51	100	101	2	103	57
	Medical Care	days	2	21	19	21	22	18	24	4
	ALL who travelled	days	18	41	2	6	35	0	317	79



Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1] (cont.)

	n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
ITEM 5								
FREQUENCY OF WATCHING ENGLISH /month	49	26	5	10	60	0	60	24
ITEM 6								
TIME WATCHING ENGLISH hrs/wk	49	7	2	4	8	0	30	9
ITEM 7								
FREQUENCY OF LISTENING TO ENGLISH /month	49	17	2	10	30	0	60	21
ITEM 8								
TIME LISTENING TO ENGLISH hrs/wk	49	3	1	2	3	0	30	6

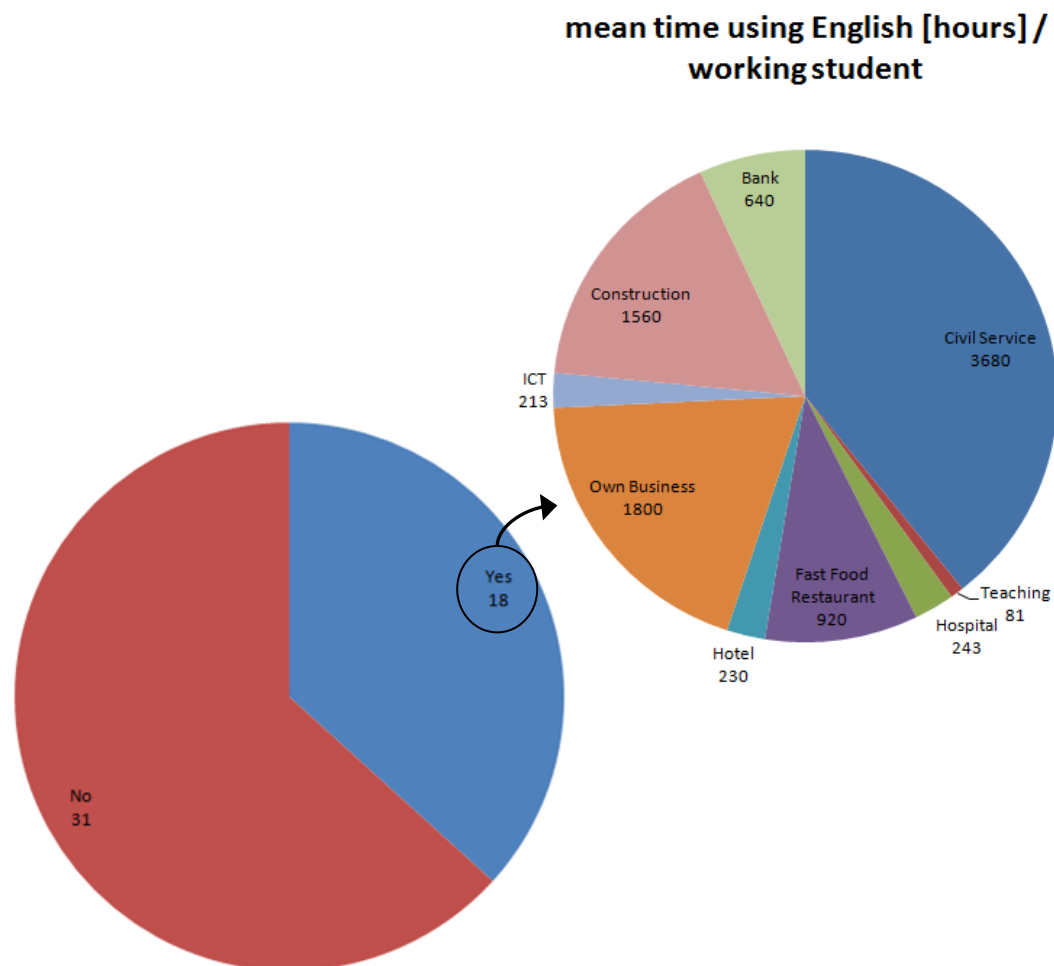


Appendix-21: Results of the Survey of Exposure to English Learning [see Section-4.5.1] (cont.)

ITEM 9

WORK EXPERIENCE FACTORED BY PROPORTION OF TIME USING ENGLISH

				n	MEAN	Q1	MEDIAN	Q3	MIN	MAX	SD
A. TYPE	Civil Service	hours		5	3680	60	1248	2520	14	14560	6168
	Teaching	hours		4	81	9	69	141	7	180	87
	Hospital	hours		4	243	145	210	308	40	513	198
	Fast Food Restaurant	hours		1	920	920	920	920	920	920	N/A
	Hotel	hours		1	230	230	230	230	230	230	N/A
	Own Business	hours		2	1800	1026	1800	2575	252	3349	2190
	ICT	hours		2	213	131	213	296	48	378	233
	Construction	hours		1	1560	1560	1560	1560	1560	1560	N/A
	Bank	hours		1	640	640	640	640	640	640	N/A
	ALL who worked	hours		18	1504	51	379	1226	7	14740	3434



Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7]

NOTE: “Xxxxx” is used to preserve anonymity

R Researcher

INTERVIEWEE-09 Xxxx Xxx XXXXXXXX XXXXXXXX

R OK. As-Salaamu alaykum. Xxxx. Thank-you very much.

INTERVIEWEE-09 Wa alaykum as-Salaam. You’re welcome.

R Thank-you very much for being happy to be interviewed right now. As you know, all this information will be very useful to my research and to my future teaching. As I also said, I will be recording the interview, so I can make notes about it.

INTERVIEWEE-09 OK, OK, no problem.

R First of all how are your current studies going? How are you doing?

INTERVIEWEE-09 Yes, it is good, yeah.

R Are you happy?

INTERVIEWEE-09 Yes. I am happy.

R Good. Your life in college now. Are you happy, or maybe you are happy to finish soon!

INTERVIEWEE-09 I happy to finish soon!

R [Laughter].

INTERVIEWEE-09 [Laughter].

R OK. Now, please Xxxx be honest. Tell me the truth. Tell me what you feel, OK. Erm. Don’t just say nice things, because you want to make your old teacher happy. OK, no. I need to know you see. Remember I am no longer teaching you now.

INTERVIEWEE-09 Yes.

R There is nothing, I have no exam with you, I don’t mark any exam of yours, OK?

INTERVIEWEE-09 OK.

R So be relaxed, you know. And also please tell me everything.

INTERVIEWEE-09 OK

R Try to think of any small detail, you know. Er and we have time free now this afternoon.

INTERVIEWEE-09 Yes.

R So don’t feel rushed, don’t feel you’re pressured, don’t feel you have to finish, like this, you know. Erm, and please remember if you don’t be honest and give me everything you feel, then really there’s no benefit for this.

INTERVIEWEE-09 Aha.

R There’s no way I can really improve this method of teaching, and there’s no real way my research can develop, without your information being frank and complete.

INTERVIEWEE-09 OK, that’s fine.

R OK, first of all, the first question, what did you think about the reading programme we did?

INTERVIEWEE-09 Very good, for me very good.

R Why?

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- INTERVIEWEE-09 As you like, if you test the grammar, for example, sometime I confused about 'my' and 'me'. I don't know how can use it. Then explain 'my' and 'me' usually confuse. But sometimes when read a story and see 'my' for 'my car' like this I talk another person, oh yes I understand the difference between this. This is good for me.
- R Right.
- INTERVIEWEE-09 I like it for me. Yes it is good.
- R Great. That's interesting. Very good. Was there any other way it was useful to you. You said for grammar it was good, but was there anything else?
- INTERVIEWEE-09 Yeh. Err. Sometimes sometime it is my problem for when you explain, and you know this is my problem, for what I use 'a' and 'the'.
- R Yes, this is difficult for any student. 'A' and 'the' is always difficult to be honest.
- p.1
- INTERVIEWEE-09 Yes. When see the story and read sometime not read. Search about word in the page 'a' and 'the' and read again because I want to understand. I think er I think I know it is difficult for me and any teacher of English if you exact this grammar 'a' and 'the'. If you write one page story, er the teacher write one page about use 'a' and 'the' the story but in the writing use more 'a' and 'the', 'a' and 'the' because it give the student reading and explain before it, it's very good for understand. In reading first and then exercise is very good to understand. Is good. Finally I like this story because you have open not select one story. You go to library Jarir and buy any story you like it. And read if you want about action, if you want er.
- R Yes, yes.
- INTERVIEWEE-09 Story, again again?
- R Action?
- INTERVIEWEE-09 Action or anything. [Laughing]. I forget. This is free for the type of story.
- R Yes.
- INTERVIEWEE-09 And you have different level. It's good. I think it's better. It's very very good.
- R Right, OK. Er, in what way did you think the reading programme was not useful? Was anything bad?
- INTERVIEWEE-09 About the story?
- R About the reading programme.
- INTERVIEWEE-09 Erm. I think no. It think it's good. I think if you each semester if you have the time enough for teacher reading is good.
- R Every semester you would like this?
- INTERVIEWEE-09 Ah yes I like this every semester, 4, 5. It change. It change. If you each doing this and listen to teacher and you have homework and writing, it's normal if you change, and reading and this is good. I like it.
- R Something different?
- INTERVIEWEE-09 Yes. It is something different.
- R OK, that's good. Could I improve this reading programme? Is there something more you like or something better?
- INTERVIEWEE-09 For reading?
- R Yes. The programme I did, is there some way it can be improved?

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- INTERVIEWEE-09 Just one step, small step. I don't think good or not good, I think is easier for if you have er website, website, er and if your student have the iPad and go to this story and reading, like library open in Appstore, and any app is free and download, is reading in the class, er it's good. But be careful, any teacher you can see the iPad, maybe whatsapp maybe playing, [laughing] but is nice.
- R Yes, yes.
- INTERVIEWEE-09 And last semester.
- R To confirm, you said the iPad is good but I must check they are working and not playing or checking emails?
- INTERVIEWEE-09 Yes, that's right.
- R Good.
- INTERVIEWEE-09 Some student don't like reading. Last semester I had a project for website, I designed a website for English, and I add maybe 20 website, for you I want to show you.
- R Please, let me see this.
- INTERVIEWEE-09 Any student who go this website and see 7 columns, different columns, grammar, 7 columns reading, 7 columns writing, and er listening. If any student click this one new page, and you can see each website has more feature.
- R Very nice.
- INTERVIEWEE-09 And website very easy for you to give homework.
- R Excellent. This was one of your projects or assignments?
- INTERVIEWEE-09 Yes, last semester.
- R Nice, very good.
- INTERVIEWEE-09 I think if you give you to show the students website, my website called "Xxxx Xxxxxx".
- R Again, it's called "Xxxxxx"?
- INTERVIEWEE-09 No, "Xxxx Xxxxxx".
- R OK, right, good.
- INTERVIEWEE-09 Check the website, yes.
- R Yes, I'll check it.
- INTERVIEWEE-09 In English, I will send it to your email and you can check it.
- R Yes, I'll check it. It's very nice.
- INTERVIEWEE-09 You can er click in website and you see ICT, information about network. I think you have tutorials, and everything is good, and this website, it has helped me for this course. And also English. You see the English. Every website you give me this paper. I think you give paper?
- R Yes, yes.
- p.2
- INTERVIEWEE-09 Maybe 2 or 3. I take nice website and put it on website. Just click. Not writing.
- R Great. Why is this better? Why do students like this more? Why do they like the internet more than just paper?
- INTERVIEWEE-09 It's easy for click. To write "h" "t" "t" "o" one mistake, if one letter wrong it's not good as website. But if you click, it's easy.
- R Yes.
- INTERVIEWEE-09 Yes, easy.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R Do you remember the time I would have you sitting down for 20 minutes just reading only? What did you think about these in-class silent reading sessions? When you were just sitting and reading. What did you think about this? Was it a good thing or a bad thing, or what do you think?
- INTERVIEWEE-09 Errr. For me I think it's good.
- R Why?
- INTERVIEWEE-09 I enjoy reading story, and sometimes for get some words, and see them and translate, and oh I save it before 1 year.
- R You have? Maa Shaa Allaah good.
- INTERVIEWEE-09 Yes, I remember for some time. Err it's good, but if you bring tea or coffee [laughing] that good. Yes.
- R That's a good idea!
- INTERVIEWEE-09 If you bring any time, I know the system in the college says no, but if you bring drink and reading then you can focus on reading
- R I never thought of this.
- INTERVIEWEE-09 Yes.
- R Good idea. You mentioned about having the iPad to read a book. How about this: what do you think about watching English videos in the iPad with English subtitles ترجمة so English with English?
- INTERVIEWEE-09 Yes, yes.
- R What do you think about this? Would students like this? Would you like this?
- INTERVIEWEE-09 It's good.
- R Better than a normal book? What do you think?
- INTERVIEWEE-09 I think you have err, you can give the student choose what he want. He want a book but also if he also want the YouTube. But it is important you also check.
- R Ah check, keep an eye on them.
- INTERVIEWEE-09 Yessss. Check they use website for reading and listening or no. This important for any teacher.
- R Right of course, keep a check on things. Also we used to do it for 20 minutes. 20 minutes. Do you think that was a good time, or too long or too short? What do you think?
- INTERVIEWEE-09 I know the system in the college if you have big book studying, I think 20 minutes is nice, which gives a break to the student and refresh the mind and reading.
- R Yes, yes.
- INTERVIEWEE-09 I think the 20 minutes is the middle. Not long, not short. It's good, very good.
- R Yes, do you remember also I used to give these reading diary sheets, and you would write down information about what you read? Do you remember this, every time? I gave the piece of paper and you would write down about what you were reading, how much time you spent, do you remember this?
- INTERVIEWEE-09 Hm, hm, yes, yes.
- R What did you think of this? Was this good? Was this useful?
- INTERVIEWEE-09 Actually useful for the teacher.
- R Ah.
- INTERVIEWEE-09 Each class, er how to put. You can put more space, I think my idea, make more space for each week for writing is good.
- R Right.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

INTERVIEWEE-09 It's my opinion, but student he has special opinion for this.
R Yes, yes.
INTERVIEWEE-09 And sometime the student, or give the student each day, and at the end get all this at the end of the week, before the weekend.
R Oh, I see. Give it to the students to fill themselves every day, and then collect them.
INTERVIEWEE-09 Yes, every day. And then each week. Each week is better.
p.3

R Right, good, OK. Erm also, we were reading for about 12 weeks during the semester, do you remember?
INTERVIEWEE-09 Hm, Yes.
R We stopped before the exams. How do you feel you were becoming during these 12 weeks? How did you change during this time?
INTERVIEWEE-09 What mean about?
R About your feeling and your reading. Did anything change?
INTERVIEWEE-09 For me?
R Yes, for you.
INTERVIEWEE-09 Er, yes, yes. Reading is more, is improve language.
R You felt you were improving? More confident?
INTERVIEWEE-09 Yes, more confident for this and enjoy. More enjoy reading, and I know about this more... I think when you give the choice for the student any type I tell you, everyone check what they like.
R Right, OK. Erm. Good. What did your friends say about, what did everyone else say about the reading? Do you remember what they were saying?
INTERVIEWEE-09 Er, as far as my friends in the class, er, some of my friend like Mr. Xxxx and Mr. Xxxxxxx because they were near?
R Yes, how were they?
INTERVIEWEE-09 Sometime they sleeping [laughing].
R Oh! [Laughing].
INTERVIEWEE-09 Er.. Mr. Xxxx like this and talk to Mr. Xxxxxx.
R But he is talking about the reading though?
INTERVIEWEE-09 Sometime about the reading, sometime another thing. Me for in the middle, because two students from Riyadh, and I am from Jeddah, and some word I don't understand what they say, but for me I focus on the reading. Yes, I focus on the reading. Sometime I ask my friend what is meaning of a word.
R So do you think that if you were here with other friends from Jeddah, maybe you would find it hard to focus?
INTERVIEWEE-09 Depend. If you have organization for studying, what time for study, what time for friends, like this in middle of week, not outside. In college, mosque and home. Some service for family.
R Hmm. Yes.
INTERVIEWEE-09 Then in weekend I go my family, and I see my friends after 10 or after 12 and enjoy. After, er, after 12 pm, saying there in home for homework or study for exam
R Good, right.
INTERVIEWEE-09 Last night, go with my friends.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R Right, but do you think if you are back in Jeddah with your family and your friends, maybe you wouldn't be able to focus so much? Maybe you have many distractions?
- INTERVIEWEE-09 Err. Maybe.
- R It depends on your organization. Again, the friends the class, did they say anything about the diary sheets?
- INTERVIEWEE-09 Diary sheets, I think yes, it's all about each day is boring for students, and writing and writing and writing. I think each week, if you writing about all this week I think it's better. Or another way give the students to go home and self do it and then give you tomorrow.
- R Yes, let them do it themselves, and give it and the end of the week. Ok. I have another main question. Give me your opinion and your feelings about reading for pleasure.
- INTERVIEWEE-09 Pleasure?
- R Yes, للمتعة , what do you think about this?
- INTERVIEWEE-09 Oh, yes. This is of course, it's good for enjoy. It's better that you give the time in class. I like it.
- R I'm saying, just yourself now, after your lessons, and even now.
- INTERVIEWEE-09 It's good. Sometime reading, but you should have time each day after pray المغرب, 1 hour you be reading is good, need organization yourself. But if you not organization it's hard.
- R But how are you now? Do you read, at home?
- p.4
- INTERVIEWEE-09 After finish semester 4 or semester 3 in this college, I take 3 story or 4 story and read, but actually just reading 2 story and 2 story not reading, because not organization! With family, with friends. And if you select, select about enjoyable the story, this is good.
- R Yes.
- INTERVIEWEE-09 Sometime, I says expensive. If you buy 35 riyals Saudi for one story.
- R Yes, yes.
- INTERVIEWEE-09 And then did not get to finish story. I think 35 is expensive story. And Jarir [the main bookshop here] is حرام you know [laughing] – it takes more money.
- R [Laughing]
- INTERVIEWEE-09 But sometime, I write in Google a story
- R Ah, good, for the internet, yes.
- INTERVIEWEE-09 And read and listening. Sometime. Not each day. Sometime, when time is free, is reading. I do it maybe last semester, last summer, I do it 2 or 3. Yes, but because reading why? Because of this college. English, more English, more English. When you finish, I don't like English class! You want some time for different, you're tired.
- R Ah, I see. You want something different. You're tired, you're fed up.
- INTERVIEWEE-09 Reading makes you remember this college. For me. Yes, really. Oh college, work, oh oh.
- R You're saying when you have too much work and too much English, you don't want to read at home because it reminds you of the English and the work you're doing the college.
- INTERVIEWEE-09 Yes, yes. I talk about the summery.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R The summer, الصيف .
- INTERVIEWEE-09 Yes, I talk about the summer. Because I live in Riyadh, but I go back to Jeddah in the summer, and I don't like to remember anything about this college. I want to enjoy. And sometime you have CFP for work in the summery for this college.
- R Yes, yes that's right. So more college work.
- INTERVIEWEE-09 Yes, more college work. In the summer you have another work. I don't like it. I write this in the feedback for this summer – not give each summery for training. Training, training, training!
- R Yes, you need some break.
- INTERVIEWEE-09 Give after finish the college 3 month, then possible for any student to become working for the same company. It's good for student. And that the student maybe go travel and learn English or take course in Saudi Arabia to improve, I think this is good. I wrote this for feedback, but I don't want is the reply.
- R That's very good. XXXX, what do you think about reading even in Arabic?
- INTERVIEWEE-09 Do you enjoy reading in Arabic?
- R A read about in Arabic.
- R What do you read?
- INTERVIEWEE-09 A subject in Arabic فن التعامل مع الأشخاص. For English, art for people, how communication.
- R Ah, interpersonal skills, communication, ah right. I understand, how to deal with people.
- INTERVIEWEE-09 Yeah, personal skills, yes. It is something for me. For me, sometime I reading but I do it another way [laughing].
- R Don't worry. This is the way of education. Sometimes you learn, but you don't practice. It's very difficult sometimes. So you do read these kinds of books?
- INTERVIEWEE-09 Yes, I read it, and first book I read some kind about psychology علم النفس.
- R Yes, psychology.
- INTERVIEWEE-09 Because when in university King Abdul Aziz I complete one years, I went to complete psychology, so it is special, I very very I like it. Sometime some friend is coming to me spoken I have solution. For this college, sometime there is problem in the class, but no one, and you are the first person to know this, they come to me and whatsapp to talk about the problem with each other, the friends in the class, and I come with some solution, because I like this psychology. And I have patience صبر patience for my friends. If you say to me bad word I have patient.
- R You have patience.
- INTERVIEWEE-09 Patience. I like this psychology.
- R It helps you to deal with people.
- INTERVIEWEE-09 It help first of all for me and for my family and for my friends. I like this very much.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R Is there anything else you read regularly, a lot?
INTERVIEWEE-09 Sometime, not book but website for er for policy? سياسة.
- R Ah, politics.
INTERVIEWEE-09 Yes, sometime reading.
- R Newspapers, you mean, الجرائد و الصحف?
INTERVIEWEE-09 Sometime newspaper, sometime about websites. But not more, not more, but it's more each day read about the news. Every day. Before sleeping I check. Arabiyah, Jazeera, I check what is around this world. I see it.
- R Great. When it comes to English, what do you think? Do you read every day?
Do you read English every day?
INTERVIEWEE-09 Actually I have Twitter, some Twitter people write in English. And sometime you see many messages, so for English not sometime reading. I have Snapshot Abu Omar.
- R Abu Omar, yes, yes.
INTERVIEWEE-09 But actually I don't open. I add it but I don't like to hear sometimes. Sometimes good it is not good. For me it is boring for English. I don't like learning English in Saudi Arabia. I want to travel outside Saudi Arabia. I hope if you talk about another person you have very very improve.
- R Yes, when you talk to other people.
INTERVIEWEE-09 But here in this college, you have problem with the student, if you say some word accent is broken you explain in presentation, it's lot for yes, it is a sure for student. I think if you travel outside Saudi Arabia you can improve myself
- R Oh, OK, this is your plan maybe?
INTERVIEWEE-09 Yes, In Shaa Allaah.
- R OK, erm, good. Are there any people you know who read a lot of things? Do have friends or family who read a lot.
INTERVIEWEE-09 Read a lot yes. My family, my father.
- R Your father.
INTERVIEWEE-09 Yes, my father. He reads a lot, and friend in here this college read, Mr. XXXXXXX in another class, its reading, yes.
- R What do people think about XXXXXXX, because he's reading a lot? What do people say?
INTERVIEWEE-09 Actually, no friends is know about this in the class.
- R Oh, they don't know.
INTERVIEWEE-09 Just me know because we travel together sometime and tell you this psychology, yeah, yeah, so he tell me he like reading, he sometime tells me.
- R How about your dad, what do his brothers, your uncles say about him reading a lot? Do they say anything?
INTERVIEWEE-09 Erm, no I think, but it's I have my father certification er not high school, not high school.
- R Oh, OK.
INTERVIEWEE-09 But, if you ask him any person, I see some talk with my father, he ask ask him what graduate from any university.
- R They think he is from a university.
INTERVIEWEE-09 Yes, they think he graduate from university. He is reading a lot.
- R What kind of things does he read?
INTERVIEWEE-09 He read about er, دين I don't know in English.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R Religion.
INTERVIEWEE-09 Ok, religion. He read more about religion.
R How about news? Does he read about news?
INTERVIEWEE-09 He read about general. He read about مجلد.
R Oh, magazines.
INTERVIEWEE-09 He read about al-Bukhaaree.
R Oh right, not magazines, you mean volumes, big volumes of books.
INTERVIEWEE-09 Yes big books. And he reads about تفسير القرآن.
R Islamic books.
INTERVIEWEE-09 Yes, all Islamic books. About, sometime, he read about history, sometime about history of Saudi Arabia, and الدولة العثمانية.
- p.6
- R What do people say about this? What do you think about this? What does your mother say. He is always reading. What do they say?
INTERVIEWEE-09 It's like it some. My uncle you don't have reading writing so they are happy.
R They are happy. They think it's good.
INTERVIEWEE-09 Yes, they are happy. And they er, my father always tell me the mind need eat the mind reading.
R Right, the food for the mind is reading.
INTERVIEWEE-09 Each time take one hour or two hour for reading
R He has told you this.
INTERVIEWEE-09 He tell this.
R Good, that's interesting, very interesting.
INTERVIEWEE-09 He says all this.
R OK Xxxx. I want to ask you as well about how you felt the reading that we did was useful to you in learning vocabulary. Is reading a book enough to improve your vocabulary?
INTERVIEWEE-09 Actually, it depend about the person.
R First of all, you though.
INTERVIEWEE-09 Ah, for me. For me erm, for remember some vocabulary and so on, maybe it's useful if you have translate to Arabic and then er keep it or save.
R Memorise.
INTERVIEWEE-09 Memorise the vocabulary. I think it improves the vocabulary.
R Did you do this? When you were reading your books?
INTERVIEWEE-09 Er, first semester yes. Sometimes, because I focus with the book of this college, because I think about the grade, about the final grade, but er, some of the word if you see the nice I like it, er, I save it.
R So Xxxx, when you were sitting in the class and I asked you to read for 20 minutes, did you write and translate and memorise words?
INTERVIEWEE-09 Er, just translate. But sometime in the lab, language lab, Sometime you save, sometimes vocabulary, you save like this. I see some word I like it, I save it.
R Great OK. Er, do you think that er that reading, just reading the stories you liked to read in class, do you think this reading is enough to improve your vocabulary?

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

INTERVIEWEE-09 Just reading, don't translate, don't? Don't save any words? For me, not more. For me, 20 or 30% in 1 years. Yes, but I don't know because I don't test. I listen to my brother and my friend study in America and British. They say if you want to improve your language read every day, read every day, but I don't do it. Because we have more work in this college. Too much. Yes, very much.

R We did try to read in class every day.

INTERVIEWEE-09 Yes and we did try to study self in home and read in home, and homework and write the project, but something every day is yes good, but another way what I like for reading for story for everything, I don't have time, so I can't read. You know the last semester, 32 hours.

R Of lessons.

INTERVIEWEE-09 Yeah. And 10 material in one semester.

R Er, 10 subjects.

INTERVIEWEE-09 Yes 10 subjects. You know the organization to have time to read is difficult. Yes, very difficult. For me I think I have a plan after graduate from college. Er, 2 month or 1 months relax. From the English, because English, English, English, Oh! Too much, I hate it. Yes, too much to save and I forget, actually I hate. I wait for me I relax 1 month to fresh the mind. I know I maybe forget some vocabulary, some of the word, but I can after 2 month study in Saudi Arabia or outside of Saudi Arabia. Travel to language centre for special for study English. I think like this.

R OK, you know you said that er, one way of learning the words was by translating, writing and then memorizing.

INTERVIEWEE-09 Yes.

R Would it help if the teacher actually tested you as well?

INTERVIEWEE-09 Test?

R Mmm. Would that help to learn more vocabulary?

INTERVIEWEE-09 Sometime. Sometime.

R For you, what do you think? Would you have liked me to test you?

p.7

INTERVIEWEE-09 Depend, depend. I think for me the smart the teacher in the college, any teacher if you have tests to improve to the students, see the schedule of the student. If you have a lot, don't give more tests.

R If the student has a lot.

INTERVIEWEE-09 Yes like this more subject. If you have for example 9 or 10 or 8 subjects in 1 semester and if you have each week test, test, test, I think not good. But you have like first semester you have more English, 1 week, or after 2 week or after 3 week, and test the student. Maybe it help.

R Every week is OK maybe in the beginning

INTERVIEWEE-09 Maybe in the beginning have free time, every student test 2 weeks, if it helps the student.

R Should that test also add to your final grade? Would that help you to learn more vocabulary?

INTERVIEWEE-09 This difficult question. Difficult question for me. Sometime is good, sometime is not good. Sometime the student do not care.

R Why don't they care?

INTERVIEWEE-09 Some student do not care! For this student you should say I give this point for bonus. I give you some grade of 10 in the final, maybe its helps.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R
INTERVIEWEE-09 As an extra.
Yes, bonus, extra. And if you, er, not bonus, but the grade, the final grade not more than 20, so each week 2 mark, 2 mark, 2, mark, 2 mark, 2 mark. In the final, for example I have 2 week for the final, if some weeks not solve the points, you have 4 points. I think maybe sometimes you have the student like this.
- R
INTERVIEWEE-09 Your friends in the class, did you see or notice anyone in the class learning a lot of vocabulary Just by reading. Only reading, not writing, not translating, not memorizing, just reading.
Actually, actually, for the student here in this college, not any student say the friend the truth.
- R
INTERVIEWEE-09 Ah.
Actually. For example, if you have exam or final exam, he see the friend, you say “you study yesterday”, he say “no”.
This is the problem.
- R
INTERVIEWEE-09 Yeah, yeah. When he have good mark, he don’t say. He says Al-Hamdulillaah OK. Or he says the bad “I take the bad mark,” but he take A+ or A!
[Laughing].
- R
INTERVIEWEE-09 Yeah [laughing] this is happen here. Because I think if you ask any friend “Do you read, you read?” they say “Oh no.” In the first semester, I see some student have good language. Then I think I ask him how to do for improve your language? He say just listen to the word then I save it. Just this. OK. I leave him for 2 week and ask again how is improve your language, he say just I watch the movie, or listen to music. Different answer.
A different answer.
- R
INTERVIEWEE-09 Yes. I know the students not all say the truth. You think if you help my friend some not all, then you have good grade for me.
He may get a better grade than you if you help him.
- R
INTERVIEWEE-09 Yeah. And this أخلاقيات is not Islamic.
Characteristics.
- R
INTERVIEWEE-09 Like this.
- R
INTERVIEWEE-09 Thank-you very much Xxxx. Thank-you so much.
That’s OK.
- R
INTERVIEWEE-09 Thanks once again Xxxx for continuing with this. Jazaakallaahu khayran. Now Xxxx, do you remember I always said that this reading has no connection with your final grade.
Yes.
- R
INTERVIEWEE-09 Nothing to do with the exams or anything. How did you feel about this?
Very, very, very good.
- R
INTERVIEWEE-09 Why, tell me why?
Because if you see any time for the grade you don’t useful, you don’t learn, you think about the grade. For me if you some teacher gives me some exercise, you see no problem mistake, but you have the grade so you forget the mistake.
- R
INTERVIEWEE-09 You forget the mistake, hm.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- INTERVIEWEE-09 Yeah. I will do it because I have the grade. This is not good. If you mistake, general mistake, and learn the mistake, it's good. Everything. If I have the grade, I think not. It's not good for reading. But if you give the students open for the reading, then that's very good.
- R Right, good, I mean, do you think I should make the reading a part of the grade?
- INTERVIEWEE-09 The grade? No. Please, please not teacher. Yes keep it separate for the student reading. Enjoy. Each point about this college is the grade. He not do it about learning. He do it about the grade. Not about the grade. It is a mistake. He the student feel the bad. And I like for this point each student read for self. If your student read high each student listen to him is not good because he is worry when reading, yes nervous, if he have some mistake for reading. Er some student listen and er يضحك love?
- R Laugh.
- INTERVIEWEE-09 Yeah, laugh. If there is some mistake word, they laugh and he feel self shy. He have for himself not good. By self better, it is better. But it is good for you I think, each class, each week self reading. I think this is very good. But if you have the grade, I think not good. In my opinion.
- R But er, what about the rest of your friends in the class? How did they feel? They know they are reading and it has no connection to the final grade. What did they say to you? Do you remember hearing them? Did they say anything?
- INTERVIEWEE-09 No I think nobody say anything to me. Maybe one person he say good no problem.
- R I see. You know erm, you know the final grade in this college, is it very very important to you?
- INTERVIEWEE-09 Yes. Final grade very very important.
- R Why?
- INTERVIEWEE-09 Because if you finish this college I want to go the company for work or XXXX, if I have low the grade I can't have any chance for working. Also for complete, if you have, for example the final grade for the student or the GPA 3.50 and you want to complete study outside Saudi Arabia, I can't.
- R You want to go outside as well.
- INTERVIEWEE-09 Yes. If you go to another company, big company, if you have low GPA, you not working nice, but if you have high, you can select anything you want: big company, small company, bank, anything you want; government, er, offices.
- R Oh, officer, you mean the army?
- INTERVIEWEE-09 Yes, army. In Saudi Arabia, army is good salary.
- R Right.
- INTERVIEWEE-09 If you have 1 star, you get 13 thousand in one month.
- R 1 star!
- INTERVIEWEE-09 Yes, and you have 2 stars and I have another certification like engineer, you have more 3000, so you take 16 thousand in a month.
- R So what grade do you need, what is a good grade? What GPA is good, for you to be able to choose easily?
- INTERVIEWEE-09 For me?
- R Hm, what GPA are you aiming at?
- INTERVIEWEE-09 But I don't have ability to take this.
- R No, you do In Shaa Allaah.
- INTERVIEWEE-09 4.5 or above excellent. It's good.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R
INTERVIEWEE-09 4.5. Right.
R But here in this college very difficult. My GPA in first semester 4.29.
R That's good.
INTERVIEWEE-09 But that is down too far. But er, 2 semester I push to 4.14 and now the last semester I grade 4.22.
R OK.
INTERVIEWEE-09 Not as good.
R Why, if you say you can get a good salary with a good grade, you are a young man and single, so do you need a big salary?
INTERVIEWEE-09 Er, yes.
R Why? What are things you need to spend on?
INTERVIEWEE-09 You need a good salary because if you want married, the cost maybe in Saudi Arabia, some friends married, the cost maybe 150000.
p.9
- R 150 thousand.
INTERVIEWEE-09 For simple, yes for simple marriage.
R Simple?
INTERVIEWEE-09 Yes, simple. For قاعة الاحتفالات.
R The wedding hall.
INTERVIEWEE-09 Yes, the wedding hall maybe 30000 is one night. For the dinner maybe also 30000, and 30000 for the wedding gift for the woman. And you need small, er, شقة .
R Apartment. Rent an apartment for 1 year at least.
INTERVIEWEE-09 Yes, you need spend like this and more. This is simple marriage.
R Do you want to do this? Do you want to marry after you finish?
INTERVIEWEE-09 Usually I told you my family I want marry, I want to marry, and last week I tell them I don't like this married.
R [Laughing].
INTERVIEWEE-09 After this college I need 1 year, 2 years for enjoy, for travel, for see like this, then look for the woman [laughing].
R OK, good. Erm, erm, anything else? This is very interesting. I also gave you in the reading a free choice. I allowed you to read anything.
INTERVIEWEE-09 Yes.
R What do you think about this?
INTERVIEWEE-09 Is good for student and good for you.
R Why, tell me why?
INTERVIEWEE-09 Good for you if you give a chance the student select the story, you can see what the student think.
R Ah, yes.
INTERVIEWEE-09 Yes. You see the student what's think. So they make decision, and see. This, I know every student and every person and every teacher each year it change the mind and the think. You give him the time if you enjoy, is reading anything.
R What was good for you though? Why did you think it was good? I gave you a free choice. Why was this good for you?
INTERVIEWEE-09 For me I choose any book I like. I like it I choose. Not the choice for the teacher. For me! And keep it for me. I can read it any time. And after finishing, give my friend or my brother. It's good.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- R Do other teachers before or now give you lots of choices when you learn and study? Or do they give you one thing?
- INTERVIEWEE-09 No. Some teacher give you different method. But only the teacher give learnt you teacher English or another subject. One thing.
- R English teachers, have there been any English teachers that tell you go and read what you like or watch what you like?
- INTERVIEWEE-09 About the reading, no. About the reading, no.
- R Anything else though? Do they give you a choice in anything?
- INTERVIEWEE-09 Sometimes they give, they say outside the college if you have book, bring some book and give another grammar for improve. Some teachers.
- R Oh, I see. They want you to bring some information about grammar from another book.
- INTERVIEWEE-09 Yeah. From my research. I see the book in this college not more grammar, so give something which has more details about that.
- R The book did not have so much grammar, so the teacher wanted you to find out more.
- INTERVIEWEE-09 Yes, yes. And then give lesson on this.
- R Was this good? Did you like this?
- INTERVIEWEE-09 Sometime. Sometime I don't like it because I know the final exam, all the student some question from this book. I need in this book, I need the grade. The exam is based on the book.
- R Yeah. I need learn this book of the college, not for learning, for the grade, actually.
- R For the grade.
- INTERVIEWEE-09 Reading and writing for the grade, the final.
- R Most of your friends are like this? They just want the grade only?
- INTERVIEWEE-09 Yes, about the grade, I think 80% for the student here think about the grade, not the learning.
- p.10
- R I see. Also, remember I said free choice of reading, but not connected to engineering, not connected to technical things. What did you think about that? Would you have preferred more books like technical books?
- INTERVIEWEE-09 If you read like technical in this class in English, and then finish the class and go another class where talk about the technical, and another class of some more technical and back to home, and solution the homework. This is not good. If you eat one food every day.
- R [Laughing]. That's a nice phrase.
- INTERVIEWEE-09 Yes, if you have free choice for the student read, I think it is better.
- R That's very interesting how you said it's like the same food all day.
- INTERVIEWEE-09 Yeah.
- R Xxxx, do you read anything about your ICT? Magazines or books or anything like this?
- INTERVIEWEE-09 Not more. Sometime reading, not English. In Arabic.
- R Why sometimes? In what situations do you read?
- INTERVIEWEE-09 I have free time. Sometime when smoke شيشة [laughing]. Reading with this.
- R You read technical things at this time?

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

INTERVIEWEE-09 Not just this time. Any time that is free. And Twitter, sometime Twitter, you have website like Windows 8. But you search about technical if you have presentation. This college you have presentation then you start reading. Searching the internet, take some information, and prepare for presentation.

R So if there's no presentation, then شيشة yes?

INTERVIEWEE-09 [Laughing].

R OK, out of all the books you've been reading, what are the best books, what are the ones you like the most?

INTERVIEWEE-09 In English, I read in English one book is very very very big this book, but not complete.

R How about in Arabic, what are the best things you read in Arabic?

INTERVIEWEE-09 In Arabic to read? فن التعامل and psychology.

R Ah yes, you said this before last time.

INTERVIEWEE-09 Yes. How to deal with people.

R Very useful. The whole world is about dealing with people.

INTERVIEWEE-09 Yes, you see when reading this information that not all the people can do it. Of course, people are not the same. Great, that's really good. I maybe have one last main question. You said you read about your technical subjects usually when you have a presentation.

INTERVIEWEE-09 Yes, usually reading when I have presentation. And presentation that is important about the grade. Prepare is good. And sometimes some word is difficult for me, accent, pronunciation. Oh.

R Yes, for technical words.

INTERVIEWEE-09 Some words are big word and this new for me. And prepare is reading sometimes 3 or 4 times. Sometime no. Write in paper, see and read.

R Why do you only read about your engineering when you have a presentation?

INTERVIEWEE-09 Because I have presentation, I have the grade.

R Don't you enjoy reading about engineering?

INTERVIEWEE-09 Well, not more.

R Why is that?

INTERVIEWEE-09 Because I study technical and read technical and too much.

R Right, too much.

INTERVIEWEE-09 Too much. But sometime sometime I like the hackers information about hackers, but I don't have time for reading. I know it's technical, each month or each week everything its change. Everything is new. I know my specialization major Networks is study this way, but when work engineer, you see another way.

R Yes, after just 1 year maybe.

INTERVIEWEE-09 Yes. You see different things. When I connect bachelors thesis from room server to student activities, fibre cable, I study about the fibre cable, but when I connect I see a new device, how it is connected, how to do it. This is new to me.

R Different to what you studied.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

- INTERVIEWEE-09 Yes. I talked with Mr. Xxxxx. I say this is new. He tell me you need at least 3 years experience for work. He tell me don't think about the salary. Think about experience. Go to any company for work. After 3 years you can learn about Networks exactly. Like this I don't like more reading. This is one reason. I like the work.
- R Work is more important than reading. Reading by itself is not beneficial. You need the experience.
- INTERVIEWEE-09 Exactly. If you read, read, read, read, and you have certification, you have course, you have the final good company, all certification, so much information, OK, now work! And see! It's difficult.
- R Maa Shaa Allaah very good. You've given me lots of excellent information. Very good. Is there anything else you want to add? Is there anything else you want to say about all this reading? Is there anything you have on your mind?
- INTERVIEWEE-09 Hmm, I told you already I have website.
- R Yes 'Xxxx Xxxxxx'. I remember from before.
- INTERVIEWEE-09 'Xxxx Xxxxxx'. If you want search normal, it is difficult to find, but I send to you website and you select and you see the website and see it like this and maybe can lead the student to this website, and click and you can see websites for English. Maybe you can make your special page for Wix. You can say Wix website.
- R Wix? What are Wix?
- INTERVIEWEE-09 Wix you can make small design page for you. And can give the student, and student can go to website and send message to website, and if you have more information, and give or put in website, and the website, any student can click it.
- R A personal website for the teacher and the students in the class.
- INTERVIEWEE-09 Yes. If some student absent, maybe go to teacher website and see all the information of the class. And you have homework in the website. If you come tomorrow, I am absent, go to website and check you have homework, I write it in website. This I think is very good.
- R Good. It's very good. Thank-you very much. You've given some very good ideas, new ideas. Very nice.
- INTERVIEWEE-09 OK, thank-you.
- R Really, thank-you for your time and comments. I am really very grateful about this and very appreciative of this. Now I wish you the best in your studies, and you finish soon and you become free! You wanted to travel?
- INTERVIEWEE-09 Yes, maybe for Indonesia or Philippines because not more money. Yes it cheaper there. Last year I go travel to Europe, Switzerland.
- R Yes, I saw that photograph on your Whatsapp. How was Switzerland, was it nice?
- INTERVIEWEE-09 Very nice.
- R Did you have chance to speak English?
- INTERVIEWEE-09 Yes, for me very good.
- R Did you go by yourself or with friends?
- INTERVIEWEE-09 With my sister, my brother-in-law, and I am speaking man. Er, they don't speak English so they said come with me to speak English. I speak in hotel and the airport, I do it everything. I speak the trip. I do it everything.

Appendix-22: A Full Transcript of one of the In-Depth Interviews (Interviewee-09) [see Section-4.7] (cont.)

R **You did everything. Maa Shaa Allaah very good.**
INTERVIEWEE-09 **I travelled together. I for my sister, I say not like sister, more like mother, because he married but don't have children, but I live with them, because for 12 or 13 years without children. I sometimes she like my mother.**

R **She is much older than you.**
INTERVIEWEE-09 **More old. I think she old 48, like this.**
R **Maa Shaa Allaah, she is older than me!**
INTERVIEWEE-09 **[Laughing].**
R **OK, Xxxx, thank-you very much. I'll stop you there. Thank-you for your time. It was excellent, really. You gave me some very interesting information.**

p.12

Appendix-23: Initial Codes generated during Phase-2 of the Interview Analysis [see Section-4.7]

STAGE-2: INITIAL CODES

Academic Studies are Important
Books Cost Money
Discipline, Scaffolding & Follow-Up are Needed to Learn
English is a Difficult language
English is Important for General Life
English is Important for Hobbies
English is Important for Work
English is Not Important for Work
English Learning Beyond Classroom
ER Culture is Lacking
ER Culture is Present
ER is Considered a Study Activity
ER is Considered Useful
ER is Important for Hobbies
ER is Important for Work
ER is Not Considered Useful
ER of Technical Material
ER Programme - Suggested Changes
ER Programme Can Not Help All Beginners
ER Programme Effects
ER Programme was Enjoyable
ER Programme was Not Enjoyable
ER Programme was Not Useful
ER Programme was Useful
Family Ties & Responsibilities are Strong
Grades are Not Important
Grades are Very Important
Hobbies, Activities & Lifestyle
Job Security
Jobs are Very Important
Learning using Technology is Attractive
Marriage is Important so Job is Important
New to Riyadh and the College
Pleasing or Impressing the Teacher
Private Business is Important so Job is Very Important
Reading Diaries Made the ER Less Interesting
Reading Diaries were Useful
Teaching Style Differs from School & Family Experiences
Vocabulary is Important
Work Load at College

Appendix-24: Main Codes & Sub-Codes generated during Phase-3 of the Interview Analysis [see Section-4.7]

MAIN CODE	SUB-CODE	Count MAIN CODES	Count SUB-CODES
ER Culture is Lacking	Abstract	94	42
ER Culture is Lacking	ER is Not Enjoyable	94	27
ER Culture is Lacking	ER is Difficult	94	11
ER Culture is Lacking	ER Lacking When Young	94	7
ER Culture is Lacking	ER of Books is Not Enjoyable	94	7
ER Programme - Suggested Changes	Vocabulary Exercises	80	11
ER Programme - Suggested Changes	SSR Sessions	80	8
ER Programme - Suggested Changes	Reading Diary Sheets Using Technology	80	7
ER Programme - Suggested Changes	Watching Videos with Subtitles	80	7
ER Programme - Suggested Changes	Browsing the Internet	80	6
ER Programme - Suggested Changes	Reading Should be Continued in Other Semesters	80	6
ER Programme - Suggested Changes	Testing the Reading	80	6
ER Programme - Suggested Changes	Writing Exercises	80	6
ER Programme - Suggested Changes	Marks Should be Given for the Reading	80	5
ER Programme - Suggested Changes	Books Selection by the Teacher	80	4
ER Programme - Suggested Changes	Abstract	80	3
ER Programme - Suggested Changes	Changes May Not be Effective	80	3
ER Programme - Suggested Changes	Discussions about the Reading	80	2
ER Programme - Suggested Changes	Reading Class Should be a Dedicated Lesson	80	2
ER Programme - Suggested Changes	Reading Class Should be Less Frequent	80	2
ER Programme - Suggested Changes	Reading Aloud	80	1
ER Programme - Suggested Changes	Reading Should be More Intensive at the Beginning	80	1
Hobbies, Activities & Lifestyle	Movies or Soap Operas	78	18
Hobbies, Activities & Lifestyle	Abstract	78	10
Hobbies, Activities & Lifestyle	Serving the Family	78	5
Hobbies, Activities & Lifestyle	Sleeping	78	5
Hobbies, Activities & Lifestyle	Social & Family Functions	78	5
Hobbies, Activities & Lifestyle	Social Media	78	5
Hobbies, Activities & Lifestyle	Travelling Abroad	78	5
Hobbies, Activities & Lifestyle	Business or Work	78	4
Hobbies, Activities & Lifestyle	Current Affairs	78	4
Hobbies, Activities & Lifestyle	Drinking Tea or Coffee Socially	78	4
Hobbies, Activities & Lifestyle	Hunting	78	4
Hobbies, Activities & Lifestyle	Cars	78	3
Hobbies, Activities & Lifestyle	Documentaries	78	2
Hobbies, Activities & Lifestyle	Browsing the Internet	78	1
Hobbies, Activities & Lifestyle	Football	78	1
Hobbies, Activities & Lifestyle	Radio	78	1
Hobbies, Activities & Lifestyle	Technology	78	1
Grades are Very Important	Abstract	77	59
Grades are Very Important	Job Selection	77	12
Grades are Very Important	Culture	77	4
Grades are Very Important	Prestige	77	2

Appendix-24: Main Codes & Sub-Codes generated during Phase-3 of the Interview

Analysis [see Section-4.7] (cont.)

Discipline, Scaffolding & Follow-Up are Needed to Learn	Exercises & Tests Based on the Reading are needed	73	20
Discipline, Scaffolding & Follow-Up are Needed to Learn	Teacher is Expected to Control the Learning	73	13
Discipline, Scaffolding & Follow-Up are Needed to Learn	Checking that the Students are Reading is needed	73	11
Discipline, Scaffolding & Follow-Up are Needed to Learn	Challenges are Faced in Different Ways	73	7
Discipline, Scaffolding & Follow-Up are Needed to Learn	Discipline is needed to Encourage Reading	73	6
Discipline, Scaffolding & Follow-Up are Needed to Learn	Exercises & Tests are needed for English Learning	73	5
Discipline, Scaffolding & Follow-Up are Needed to Learn	Vocabulary Tests	73	5
Discipline, Scaffolding & Follow-Up are Needed to Learn	Lack of Discipline is Not Respected	73	3
Discipline, Scaffolding & Follow-Up are Needed to Learn	Book Summaries	73	1
Discipline, Scaffolding & Follow-Up are Needed to Learn	Discipline is a Part of the Culture	73	1
Discipline, Scaffolding & Follow-Up are Needed to Learn	Maturity is Lacking in Many Students	73	1
ER Culture is Present	Abstract	54	8
ER Culture is Present	Newspapers & Magazines	54	8
ER Culture is Present	Older People Read	54	8
ER Culture is Present	ER is Enjoyable	54	7
ER Culture is Present	Religious Literature	54	5
ER Culture is Present	Videos with Subtitles	54	5
ER Culture is Present	ER is Not Difficult	54	3
ER Culture is Present	Social Media	54	3
ER Culture is Present	Comics	54	2
ER Culture is Present	Websites in General	54	2
ER Culture is Present	History	54	1
ER Culture is Present	Novels	54	1
ER Culture is Present	Poetry	54	1
Jobs are Very Important	Abstract	42	16
Jobs are Very Important	Serving the Family	42	10
Jobs are Very Important	Salary	42	5
Jobs are Very Important	Cars	42	3
Jobs are Very Important	Unemployment	42	3
Jobs are Very Important	Prestige	42	2
Jobs are Very Important	Social & Family Functions	42	2
Jobs are Very Important	Culture	42	1
Learning using Technology is Attractive	Internet	42	20
Learning using Technology is Attractive	Mobile Phones & Tablets	42	19
Learning using Technology is Attractive	Laptops	42	3
English Learning Beyond Classroom	Using Technology	41	13
English Learning Beyond Classroom	Reading	41	10
English Learning Beyond Classroom	Travelling Abroad	41	4
English Learning Beyond Classroom	Improving Professional Performance	41	3
English Learning Beyond Classroom	Watching Videos	41	3
English Learning Beyond Classroom	Grammar	41	2
English Learning Beyond Classroom	Vocabulary	41	2
English Learning Beyond Classroom	Exercises	41	1
English Learning Beyond Classroom	Improving Academic Performance	41	1
English Learning Beyond Classroom	Listening	41	1
English Learning Beyond Classroom	Speaking	41	1

Appendix-24: Main Codes & Sub-Codes generated during Phase-3 of the Interview Analysis [see Section-4.7] (cont.)

MAIN CODE	SUB-CODE	Count MAIN CODES	Count SUB-CODES
Teaching Style Differs from School & Family Experiences	Lessons were Too Many or Too Difficult	39	7
Teaching Style Differs from School & Family Experiences	Confidence is Lacking	39	5
Teaching Style Differs from School & Family Experiences	Father is Distant & Strict	39	5
Teaching Style Differs from School & Family Experiences	Backgrounds & Cultures Differ	39	4
Teaching Style Differs from School & Family Experiences	Teacher is Distant & Strict	39	4
Teaching Style Differs from School & Family Experiences	Teacher Usually Checks on Students' Work	39	4
Teaching Style Differs from School & Family Experiences	Abstract	39	2
Teaching Style Differs from School & Family Experiences	Few Choices	39	2
Teaching Style Differs from School & Family Experiences	Learning Facts	39	2
Teaching Style Differs from School & Family Experiences	Problem Solving Practice is Lacking	39	2
Teaching Style Differs from School & Family Experiences	Praise	39	1
Teaching Style Differs from School & Family Experiences	Textbooks Were Central to Learning	39	1
Family Ties & Responsibilities are Strong	Serving the Family	36	13
Family Ties & Responsibilities are Strong	Staying with the Family	36	13
Family Ties & Responsibilities are Strong	Marrried Students feel More Responsible	36	4
Family Ties & Responsibilities are Strong	Prestige	36	4
Family Ties & Responsibilities are Strong	Abstract	36	2
Work Load at College	Lessons were Too Many or Too Difficult	34	22
Work Load at College	ER Enjoyment Reduced Because of Work Load	34	6
Work Load at College	ER Time Reduced Because of Work Load	34	4
Work Load at College	Abstract	34	1
Work Load at College	Praise	34	1
Vocabulary is Important	Memorising Words	30	10
Vocabulary is Important	Translating Words	30	10
Vocabulary is Important	Abstract	30	4
Vocabulary is Important	Writing Words	30	4
Vocabulary is Important	Listening to Words	30	2
ER is Considered Useful	Abstract	27	8
ER is Considered Useful	College Assignments	27	4
ER is Considered Useful	Writing	27	4
ER is Considered Useful	Vocabulary Learning	27	3
ER is Considered Useful	Exercising the Mind	27	2
ER is Considered Useful	Grammar	27	2
ER is Considered Useful	Learning in General	27	2
ER is Considered Useful	Overall English	27	1
ER is Considered Useful	Speaking	27	1
Marriage is Important so Job is Important	Abstract	22	10
Marriage is Important so Job is Important	Marrying Relatively Young	22	8
Marriage is Important so Job is Important	Cost of Getting Marrried	22	4

Appendix-24: Main Codes & Sub-Codes generated during Phase-3 of the Interview

Analysis [see Section-4.7] (cont.)

New to Riyadh and the College		21	0
English is a Difficult language	Reading is Difficult	19	5
English is a Difficult language	Reading is Difficult for New Students	19	5
English is a Difficult language	Abstract	19	4
English is a Difficult language	English is Difficult for New Students	19	3
English is a Difficult language	Challenges are Faced in Different Ways	19	2
ER is Considered a Study Activity	College Assignments	19	4
ER is Considered a Study Activity	Reading is Associated with School Studies	19	4
ER is Considered a Study Activity	Books are to be Studied	19	3
ER is Considered a Study Activity	Vocabulary Learning	19	3
ER is Considered a Study Activity	Grammar	19	2
ER is Considered a Study Activity	Overall English	19	2
ER is Considered a Study Activity	Writing	19	1
ER of Technical Material	ER of Technical Material is Not Considered Useful	16	12
ER of Technical Material	ER of Technical Material is Considered Useful	16	4
ER Programme Effects		10	0
Pleasing or Impressing the Teacher		10	0
ER Programme was Useful		9	0
English is Not Important for Work		8	0
ER Programme was Enjoyable		7	0
Academic Studies are Important	Reading Only for Academic Studies	6	4
Academic Studies are Important	Desire to Improve Oneself	6	2
Books Cost Money		6	0
ER is Not Considered Useful		6	0
English is Important for Work		5	0
ER Programme Can Not Help All Beginners		5	0
English is Important for General Life		4	0
ER Programme was Not Enjoyable		4	0
Reading Diaries Made the ER Less Interesting		4	0
ER is Important for Hobbies		2	0
ER is Important for Work		2	0
ER Programme was Not Useful		2	0
Grades are Not Important		2	0
Private Business is Important so Job is Very Important		2	0
Reading Diaries were Useful		2	0
English is Important for Hobbies		1	0
Job Security		1	0

Appendix-25: Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7]

[illegible]

Appendix-25: Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7] (cont.)

2. Why are grades so important?	Grades are Very Important	Grades are Very Important Grades are Very Important Grades are Very Important Jobs are Very Important Grades are Very Important Jobs are Very Important Grades are Very Important Jobs are Very Important Family Ties & Responsibilities are Strong Jobs are Very Important Job Security	Abstract Culture Prestige Abstract Job Selection Culture Prestige Prestige Prestige Unemployment
	Good Grades needed for a Good Job		
	A Very Good Salary is needed	Jobs are Very Important Family Ties & Responsibilities are Strong Jobs are Very Important Family Ties & Responsibilities are Strong Family Ties & Responsibilities are Strong Marriage is Important so Job is Important Marriage is Important so Job is Important Marriage is Important so Job is Important Family Ties & Responsibilities are Strong Jobs are Very Important Jobs are Very Important Private Business is Important so Job is Very Important	Salary Abstract Serving the Family Serving the Family Staying with the Family Abstract Marrying Relatively Young Cost of Getting Married Married Students feel More Responsible Social & Family Functions Cars

Appendix-25: Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7] (cont.)

NARRATIVE	PHASE-4 POTENTIAL THEME	PHASE-3 MAIN CODE (or SUB-CODE)	PHASE-3 MAIN CODE (or SUB-CODE)	
3. Open to ER if done differently	Some ER Culture is Present	ER Culture is Present	Abstract	
		ER Culture is Present	Older People Read	
		ER Culture is Present	ER is Enjoyable	
		ER Culture is Present	ER is Not Difficult	
		ER Culture is Present	Newspapers & Magazines	
		ER Culture is Present	Religious Literature	
		ER Culture is Present	Poetry	
		ER Culture is Present	History	
		ER Culture is Present	Videos with Subtitles	
		ER Culture is Present	Social Media	
		ER Culture is Present	Websites in General	
		Learning using Technology is Attractive	Internet	
		Learning using Technology is Attractive	Mobile Phones & Tablets	
		Learning using Technology is Attractive	Laptops	
		ER Culture is Present	Comics	
		ER Culture is Present	Novels	
		ER is Considered Useful	ER is Considered Useful	Abstract
			ER is Considered Useful	Learning in General
			ER is Considered Useful	Exercising the Mind
			ER is Considered Useful	Overall English
	ER is Considered Useful		College Assignments	
	ER is Considered Useful		Writing	
	ER of Technical Material		ER of Technical Material is Considered Useful	
	ER is Important for Work			
	ER is Considered Useful		Vocabulary Learning	
	ER is Considered Useful		Grammar	
	ER is Considered Useful		Speaking	
	Desire to Improve Oneself	Desire to Improve Oneself	ER is Important for Hobbies	
			English is Important for Work	
			English is Important for General Life	
			English is Important for Hobbies	
ER Programme Effects				
ER Programme was Useful				
ER Programme was Enjoyable				
Reading Diaries were Useful				
Academic Studies are Important				
Grades are Not Important				
Pleasing or Impressing the Teacher				
English Learning Beyond Classroom			Improving Academic Performance	
English Learning Beyond Classroom			Improving Professional Performance	
English Learning Beyond Classroom			Using Technology	
English Learning Beyond Classroom			Reading	
English Learning Beyond Classroom			Travelling Abroad	
English Learning Beyond Classroom			Watching Videos	
English Learning Beyond Classroom			Grammar	
English Learning Beyond Classroom			Vocabulary	
Desire to Improve Oneself			Desire to Improve Oneself	Vocabulary is Important
	Vocabulary is Important	Memorising Words		
	Vocabulary is Important	Translating Words		
	Vocabulary is Important	Writing Words		
	Vocabulary is Important	Listening to Words		
	English Learning Beyond Classroom	Exercises		
	English Learning Beyond Classroom	Listening		
	English Learning Beyond Classroom	Speaking		
	English Learning Beyond Classroom			
	English Learning Beyond Classroom			
	English Learning Beyond Classroom			

Appendix-25: Narratives and Potential Themes identified during Phase-4 of the Interview Analysis [see Section-4.7] (cont.)

4. Changes to future ER	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Abstract
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Changes May Not be Effective
	Discipline, Scaffolding & Follow-Up are Needed to Learn	ER Programme - Suggested Changes	Exercises & Tests Based on the Reading are needed
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Testing the Reading
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Marks Should be Given for the Reading
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Vocabulary Exercises
	Discipline, Scaffolding & Follow-Up are Needed to Learn	ER Programme - Suggested Changes	Vocabulary Tests
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	SSR Sessions
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Should be Continued in Other Semesters
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Class Should be a Dedicated Lesson
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Class Should be Less Frequent
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Should be More Intensive at the Beginning
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Watching Videos with Subtitles
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Browsing the Internet
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Books Selection by the Teacher
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Writing Exercises
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Discussions about the Reading
	Discipline, Scaffolding & Follow-Up are Needed to Learn	ER Programme - Suggested Changes	Book Summaries
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Aloud
	ER Programme - Suggested Changes	ER Programme - Suggested Changes	Reading Diary Sheets Using Technology

NARRATIVE	PHASE-4 POTENTIAL THEME	No. of EXTRACTS according to INTERVIEWEE												TOTAL
		01	02	03	04	05	06	07	08	09	10	11	12	
1. Problems	ER Culture is Lacking	7	5	2	4	18	15	30	12	17	14	3	26	153
	The ER Learning Style was Different to Prior Experiences	6	3	4	2	12	34	20	13	22	19	3	6	144
	The College Life Prevented ER	2	1	0	1	30	11	12	9	17	7	0	5	95
	The General Life Prevented ER	2	1	0	1	12	4	7	7	19	11	0	20	84
2. Why are grades so important?	Grades are Very Important	5	2	3	2	8	2	1	3	15	17	4	3	65
	Good Grades needed for a Good Job	2	1	2	3	8	0	0	0	5	15	4	1	41
	A Very Good Salary is needed	5	2	1	4	7	0	0	0	7	30	6	14	76
3. Open to ER if done differently	Some ER Culture is Present	6	3	3	2	6	11	10	6	19	8	3	19	96
	ER is Considered Useful	1	2	2	2	9	0	5	5	18	8	5	16	73
	Desire to Improve Oneself	10	0	5	5	5	14	9	5	18	7	2	6	86
4. Changes to future ER	ER Programme - Suggested Changes	3	4	5	7	15	21	17	6	12	11	0	5	106

Appendix-26: Narratives and Themes refined during Phase-5 & Phase-6 of the Interview Analysis [see Section-4.7]

PHASE-5 NARRATIVE	PHASE-5 THEME	PHASE-4 NARRATIVE	PHASE-4 POTENTIAL THEME
1. What could have prevented vocabulary gain during the ER programme?	The ER Culture is Lacking The ER Learning Style was Different to Prior Experiences The College Life Eclipsed ER The General Life Eclipsed ER	1. Problems	ER Culture is Lacking The ER Learning Style was Different to Prior Experiences The College Life Prevented ER The General Life Prevented ER
2. Extrinsically motivated students	Grades are Very Important Good Grades are needed for a Good Job A Very Good Salary is needed	2. Why are grades so important?	Grades are Very Important Good Grades needed for a Good Job A Very Good Salary is needed
3. ER can be more successful if administered differently	Some ER Culture is Present ER was Considered Useful Evidence of Intrinsic Motivation Proposed Changes for future ER Programmes	3. Open to ER If done differently	Some ER Culture is Present ER is Considered Useful Desire to Improve Oneself
		4. Changes to future ER	ER Programme - Suggested Changes
PHASE-6 NARRATIVE	PHASE-6 THEME	PHASE-5 NARRATIVE	PHASE-5 THEME
(A) Possible Factors for the Modest Vocabulary Gains During the ER Programme	(A1) Lack of ER Culture (A2) Conflict between the ER Learning Style and Prior Experiences (A3) College Life Inhibited ER (A4) General Life Inhibited ER	1. What could have prevented vocabulary gain during the ER programme?	The ER Culture is Lacking The ER Learning Style was Different to Prior Experiences The College Life Eclipsed ER The General Life Eclipsed ER
(B) Extrinsically Motivated Students	(B1) Strong Motivation for High Grades (B2) High Grades are Needed for Good Jobs (B3) Good Jobs Should Have High Salaries	2. Extrinsically motivated students	Grades are Very Important Good Grades are needed for a Good Job A Very Good Salary is needed
(C) Future Implementation of ER	(C1) Presence of Some Limited ER Culture (C2) ER was Considered Useful (C3) Presence of Some Limited Intrinsic Motivation (C4) Proposals for Future ER Programmes	3. ER can be more successful if administered differently	Some ER Culture is Present ER was Considered Useful Evidence of Intrinsic Motivation Proposed Changes for future ER Programmes

NARRATIVE	THEME	No. of EXTRACTS according to INTERVIEWEE												TOTAL
		01	02	03	04	05	06	07	08	09	10	11	12	
(A) Possible Factors for the Modest Vocabulary Gains During the ER Programme	(A1) Lack of ER Culture	7	5	2	4	18	15	30	12	17	14	3	26	153
	(A2) Conflict between the ER Learning Style and Prior	6	3	4	2	12	34	20	13	22	19	3	6	144
	(A3) College Life Inhibited ER	2	1	0	1	30	11	12	9	17	7	0	5	95
	(A4) General Life Inhibited ER	2	1	0	1	12	4	7	7	19	11	0	20	84
(B) Extrinsically Motivated Students	(B1) Strong Motivation for High Grades	5	2	3	2	8	2	1	3	15	17	4	3	65
	(B2) High Grades are Needed for Good Jobs	2	1	2	3	8	0	0	0	5	15	4	1	41
	(B3) Good Jobs Should Have High Salaries	5	2	1	4	7	0	0	0	7	30	6	14	76
(C) Future Implementation of ER	(C1) Presence of Some Limited ER Culture	6	3	3	2	6	11	10	6	19	8	3	19	96
	(C2) ER was Considered Useful	1	2	2	2	9	0	5	5	18	8	5	16	73
	(C3) Presence of Some Limited Intrinsic Motivation	10	0	5	5	5	14	9	5	18	7	2	6	86
	(C4) Proposals for Future ER Programmes	3	4	5	7	15	21	17	6	12	11	0	5	106